

# IEA 1958-2008: 50 years of experiences and memories

VOLUME 1

Robitaille  
Schwille  
Pelgrum  
Torney-Purta  
Takala  
Ferrer  
Masters  
Fraillon  
Schulz  
Loveless  
Brassói  
Leung  
Travers  
Brusselman - Dehairs  
Leimu  
Keeves  
Ambach  
Gubler  
Lietz  
Čepić  
Covacevich  
Japelj-Pavešič  
Ortiz  
Kádár-Fülöp  
Ainley  
Lokan  
Tiana  
Lockheed  
Postlethwaite  
Wagemaker  
Papanastasiou  
Covacevich  
Kangro  
Lagos  
Justin  
Dukynaitė  
Wolf  
Cariola  
Montie  
Plomp  
Howie  
Blahova  
Purves  
Suter  
Thomson

Edited By:

Constantinos Papanastasiou • Tjeerd Plomp • Elena C. Papanastasiou

# IEA 1958-2008: 50 years of experiences and memories

VOLUME 2

Robitaille  
Schwille  
Pelgrum  
Torney-Purta  
Takala  
Schulz  
Loveless  
Ferrer  
Masters  
Fraillon  
Brassói  
Leung  
Travers  
Brusselman - Dehairs  
Leimu  
Keeves  
Ambach  
Gubler  
Lietz  
Čepič  
Covacevich  
Japelj-Pavešič  
Ortiz  
Kádár-Fülöp  
Ainley  
Lokan  
Tiana  
Lockheed  
Postlethwaite  
Wagemaker  
Lehmann  
Papanastasiou  
Kangro  
Lagos  
Justin  
Dukynaitė  
Wolf  
Cariola  
Montie  
Blahova  
Justin  
Purves  
Plomp  
Howie  
Suter  
Thomson

Edited By:

Constantinos Papanastasiou • Tjeerd Plomp • Elena C. Papanastasiou

IEA, the International Association for the Evaluation of Educational Achievement, conducts worldwide comparative studies of educational achievement. It is well known for its cycles of studies, such as the Trends in Mathematics and Science Studies (TIMSS) and the Progress in Reading Literacy Studies (PIRLS), as well as studies focused on particular subjects and themes such as civics education and ICT in education.

IEA's founding meeting took place in 1958 at the Unesco Office in Hamburg (Germany). This first meeting, it turned out, heralded the beginning of many impressive international comparative assessment studies. Over the 50 years since that initial meeting, the IEA has developed into a highly renowned research organization with a membership of more than 60 education systems (or countries) worldwide. In 2008, the IEA celebrated its 50th anniversary at the 49th General Assembly meeting in Berlin where, during a special Roundtable Session, a number of honorary members shared their past experiences in IEA with the IEA representatives present. It was clear that the IEA had evolved over the years to become not only a vibrant, global organization, but also an international community of friends and colleagues, all of whom were deeply committed to their involvement in IEA.

This book presents a mosaic of experiences and memories of IEA members from all over the world, reflecting a wide variety of foci on the IEA and its studies, including: personal memories; how studies were originated, designed and developed; the impact of IEA studies on a country's policy and practice; and perspectives of observers involved in IEA's activities.

The 29 chapters of this book present the reader with a rich collection of memories and experiences. They reveal how relevant and important the IEA has been for educational research, policies and practices, but they also demonstrate how the IEA has, over the past 50 years, developed into a 'global community' of professionals and friends.



# IEA 1958-2008: 50 Years of Experiences and Memories



*Volume 1*



# **IEA 1958-2008: 50 Years of Experiences and Memories**

EDITED BY:

Constantinos Papanastasiou

Tjeerd Plomp

Elena C. Papanastasiou

*Volume 1*



**The International Association for the  
Evaluation of Educational Achievement  
(IEA)**

Copyright © 2011 The International Association for the Evaluation of Educational Achievement (IEA). All rights reserved. No part of this book may be reproduced in any form, by any means, including, but not limited to, the process of scanning and digitization, or stored in a database or retrieval system, without the prior written permission of the IEA.

International Association for the  
Evaluation of Educational Achievement  
Stichting IEA Secretariaat Nederland  
Amsterdam, The Netherlands  
Herengracht 487, 1017 BT Amsterdam, The Netherlands  
Phone +31 (0) 20 625 3625  
Fax +31 (0) 20 420 7136  
Email [secretariat@IEA.nl](mailto:secretariat@IEA.nl)  
[www.iea.nl](http://www.iea.nl)

Edited by  
Constantinos Papanastasiou  
Tjeerd Plomp  
Elena C. Papanastasiou

Published originally by  
Cultural Center of the Kykkos Monastery  
PO Box 28192, 2091 Nicosia, Cyprus.

Typeset: Maria Nicolaou-Prodrômou  
Printer: Kailas Printers and Lithographers Ltd  
Cover Design: Stelios Smyrillis

ISBN 978-9963-645-77-0

IEA, the International Association for the Evaluation of Educational Achievement, conducts worldwide comparative studies of educational achievement. It is well known for its cycles of studies, such as the Trends in Mathematics and Science Studies (TIMSS) and the Progress in Reading Literacy Studies (PIRLS), as well as studies focused on particular subjects and themes such as civics education and ICT in education.

IEA's founding meeting took place in 1958 at the Unesco Office in Hamburg (Germany). This first meeting, it turned out, heralded the beginning of many impressive international comparative assessment studies. Over the 50 years since that initial meeting, the IEA has developed into a highly renowned research organization with a membership of more than 60 education systems (or countries) worldwide.

In 2008, the IEA celebrated its 50th anniversary at the 49th General Assembly meeting in Berlin where, during a special Roundtable Session, a number of honorary members shared their past experiences in IEA with the IEA representatives present. It was clear that the IEA had evolved over the years to become not only a vibrant, global organization, but also an international community of friends and colleagues, all of whom were deeply committed to their involvement in IEA.

This book presents a mosaic of experiences and memories of IEA members from all over the world, reflecting a wide variety of foci on the IEA and its studies, including: personal memories; how studies were originated, designed and developed; the impact of IEA studies on a country's policy and practice; and perspectives of observers involved in IEA's activities.

The 29 chapters of this book present the reader with a rich collection of memories and experiences. They reveal how relevant and important the IEA has been for educational research, policies and practices, but they also demonstrate how the IEA has, over the past 50 years, developed into a 'global community' of professionals and friends.





# Contents

Preface .....	xiii
---------------	------

## *Volume 1*

### **PART A:**

#### **EARLY DEVELOPMENTS - HISTORY AND STUDIES**

1. IEA - From the Beginning in 1958 to 1990 <i>John P. Keeves</i> .....	3
2. IEA from 1988 – 1994: An Organization in Transition <i>Tjeerd Plomp</i> .....	41
3. The Second International Mathematics Study (SIMS): Intention, Implementation, Attainment <i>Kenneth J. Travers</i> .....	73
4. The IEA Studies in Civic Education: Their History from the Late 1960s to 2005 <i>Judith Torney-Purta, John Schwille</i> .....	97
5. The International Study of Writing <i>Sauli Takala</i> .....	115
6. IEA Assessments of Information and Communication Technologies (ICT) in Education <i>Willem J. Pelgrum, Tjeerd Plomp</i> .....	137
7. The Growth and Development of the Preprimary Project <i>Jeanne E. Montie</i> .....	165
8. The International Civic and Citizenship Education Study (ICCS) <i>Wolfram Schulz, Julian Fraillon, John Ainley</i> .....	187

9. IEA International Research Conferences (IRC) from 2004 to 2008  
*Constantinos Papanastasiou, Elena C. Papanastasiou* .....201

## **PART B:**

### **THE CONTRIBUTION OF IEA TO EDUCATIONAL RESEARCH**

10. The Relationship of IEA to some Developments  
 in Educational Research Methodology and Measurement  
 during the Years from 1962 to 1992  
*John P. Keeves, Petra Lietz* .....217
11. IEA: International Studies, Impact and Transition  
*Hans Wagemaker* .....253
12. The Contribution of IEA to Research and Education  
*Richard M. Wolf*.....273

## **PART C:**

### **REFLECTIONS ON CONTRIBUTIONS OF IEA RESEARCH FROM THE PERSPECTIVE OF EDUCATIONAL SYSTEMS**

13. The Involvement of African Countries  
 in the IEA Studies over 50 Years  
*Sarah Howie* .....289
14. The Contribution of IEA Research Studies to Australian Education  
*John Ainley, John Keeves, Jan Lokan, Petra Lietz,  
 Geoff Masters, Sue Thomson* .....317
15. Chilean Participation in IEA Studies  
*Leonor Cariola, Catalina Covacevich,  
 Johanna Gubler, Ema Lagos, Marcela Ortiz* .....373
16. The Significance of IEA Studies for Education in East Asia  
*Frederick K.S. Leung* .....389
17. The Impact of IEA on Educational Policy Making in Germany  
*Rainer Lehmann* .....411
18. How IEA Influenced the Education System in Hungary  
*Sándor Brassófi, Judith Kádár-Fülöp* .....431

19. IEA Experiences from Latvia, Lithuania,  
Slovak Republic and Slovenia  
*Hans Pelgrum, Viera Blahova, Rita Dukynaitė,  
Barbara Japelj, Andris Kangro* .....447
20. The IEA of 1990-2001: a 50th Year Reflection  
on Transformation and US Participation  
*Gordon Ambach* .....469

## *Volume 2*

### **PART D:**

#### **MEMORIES OF KEY PERSONS IN THE IEA**

21. Roundtable at the 49th General Assembly, Berlin 2008  
Christiane Brusselman- Dehairs, Tjeerd Plomp,  
Neville Postlethwaite, David Robitaille,  
Judith Torney- Purta, Alejandro Tiana, Tom Loveless  
*Edited by Judith Torney- Purta* .....493
22. The Scientific Contributions of Torsten Husén  
and Neville Postlethwaite to the Development of International  
Comparative Research on Educational Achievement  
*Rainer Lehmann* .....515
23. The Evolution of the IEA: A Memoir  
*Alan C. Purves* .....531
24. Experiencing the World as an Educational Laboratory  
*Alejandro Tiana Ferrer* .....557
25. 50 Years of IEA: Some Reflections about its History  
*Judith Torney-Purta* .....577
26. 50 years of IEA: A Personal Account  
*Kimmo Leimu* .....591
27. Experiencing Innovation and Capacity Building  
in IEA Research - A Memoir, 1963-2008  
*John Schwillie* .....627

**PART E:****MEMORIES OF FRIENDS OF THE IEA**

28. Reflections on IEA from the Perspective of a World Bank Official  
*Marlaine E. Lockheed* .....711
29. Guiding International Comparative Studies (1981 to 2009):  
Reflections of a U.S. Program Director  
*Larry E. Suter*.....721

**APPENDIX:**

- Author Biographies.....745

## Preface

In 2008, the IEA celebrated its 50th anniversary at the General Assembly (GA) meeting in Berlin. The meeting celebrated this special anniversary with a very festive roundtable session, moderated by Tom Loveless, the General Assembly Member representing the USA. Participants in this Roundtable were: Christiane Brusselman-Dehairs (Belgium), Tjeerd Plomp (The Netherlands), Neville Postlethwaite (England), David Robitaille (Canada), Alejandro Tiana (Spain), Judith Torney-Purta (USA) and they shared with the GA representatives present at the meeting some of their memories and experiences in IEA. With anecdotes, stories, personal insights and experiences, they gave a vivid picture of IEA's rich history.

Constantinos Papanastasiou, the GA representative for Cyprus, was especially intrigued by the information that was presented at the roundtable, and he felt that it was important to preserve these memories of the establishment and development of the organization, and of the people who were so active in its history. At the dinner that followed the roundtable, Constantinos suggested to Neville Postlethwaite that they could perhaps put together a book of IEA memories and experiences. Neville agreed with the idea and was very eager to be part of this process. However, in 2009, Neville Postlethwaite passed away, as did Torsten Husen who was Neville's predecessor and a founding father of the IEA.

The passing away of two of the 'giants' of IEA made Constantinos realize that we would now regularly lose people who had played key roles in the IEA both internationally and nationally; but more than that, we

would also lose their experiences, memories and stories of the IEA, the studies, the many meetings, the personalities. It was this realization that made Constantinos decide to immediately embark on the project of which this book is the concrete result. Constantinos's discussions with John Schwille at the time were vital in relation to the vision of this book, and John's input is greatly appreciated.

In the year following the 49th GA meeting in Berlin, Constantinos was able to secure funding from the Cultural Center of the Kykkos Monastery in Cyprus to prepare this book, in which key persons in IEA's history--such as GA representatives, International Study Coordinators, National Research Coordinators of studies and/or observers-- present their experiences and memorable events during the last 50 years of IEA. Thus at the subsequent GA meeting, in 2009 in Tallinn, Constantinos invited a number of people to contribute to the book, receiving many positive reactions. Because authors had quite 'free rein' regarding the topic of their paper, the papers--now chapters in the book-- offer a rich mosaic with a variety of foci, such as personal memories; how studies were originated, designed and developed; impact of IEA studies on a country's policy and practice, and/or the perspectives of observers involved in IEA's activities. This book, comprising two volumes, does not intend to present an 'authorized history' of the IEA. Instead, like a mosaic, it consists of a varied collection of personal perspectives and experiences which we would like to share with those involved or interested in the IEA.

From the first days of this undertaking, Constantinos was in frequent contact with Tjeerd Plomp, who offered significant input in relation to the book. This collaboration led to a drastically different approach to the book than Constantinos had originally envisioned, but which he felt would be so useful and significant, that in 2010, he invited Tjeerd to join him as a co-editor of this book. Throughout the process of preparing this book, Elena Papanastasiou was also invited to join as an editor to assist Constantinos in the major endeavour of preparing this book. All three editors, and especially Elena, paid special attention to

ensure the readability of each chapter for all readers, but especially for those who were themselves not actively involved in the early days of IEA, nor familiar with the contexts referred. In a few cases, we asked other people to review a chapter, and in this respect we want to express our thanks to Jan-Erik Gustafsson and Robert Garden for their help in the review of a few chapters for which we felt we needed assistance. We are also very grateful to Judith Torney-Purta for the final editing of the chapter on the Roundtable at the 49th General Assembly (Berlin, 2008).

Eventually, the many enthusiastic reactions to Constantinos's invitation resulted in this two-volume, five-part book:

- A. *Early developments*: a series of chapters on the early years of IEA, on the various studies, on the international research conferences, and on the period of transition (the late 1980s and early 1990s);
- B. *IEA and educational research*: a few chapters on IEA's contribution to research, methodology and education;
- C. *Reflections on contributions of IEA research from the perspective of educational systems across the world*: chapters discussing IEA studies by authors from a number of countries all over the world;
- D. *Memories of key persons in the IEA*: a number of chapters written by people who played a key role in the IEA, but also chapters about some key IEA people like Torsten Husen and Neville Postlethwaite, as well as what had been presented in the Round Table at the 49th GA meeting;
- E. *Memories of friends of the IEA*: two chapters by individuals involved in the IEA as observers, but not as GA representatives and/or National Research Coordinators.

The 29 chapters of this book present the reader with a rich collection of memories and experiences. They reveal how relevant and important the IEA has been for educational research, policies and practices, but they also demonstrate how the IEA has, over the past 50 years, devel-



oped into a 'global community' of professionals and friends.

We would like to sincerely thank all those who contributed chapters to this book, for their generosity and willingness to be part of this process, as well as for the timely manner in which they acknowledged our various requests. We would also like to express our appreciation to Kathy Stephanides for her skilled and careful language editing of the chapters and to Maria Nicolaou for her expertise in preparing the manuscript for printing.

Above all, we would like to especially acknowledge and thank the **Reverend Bishop Nikiforos of Kykkos and Tylliria**, who, as Chair of the Board of the Cultural Center of the Kykkos Monastery, has funded, and selected this book to be included in the publications of the Center.

We hope that the readers of this book will enjoy reading the experiences and memories of 50 years of IEA as we did!

*Constantinos Papanastasiou*

*Tjeerd Plomp*

*Elena C. Papanastasiou*

*September 1, 2011*



PART A

**EARLY DEVELOPMENTS -  
HISTORY AND STUDIES**



## CHAPTER 1

# IEA - From the Beginning in 1958 to 1990

*John P. Keeves*

### Introduction

The decade following the end of the Second World War was a period of remarkable growth in primary and secondary education across most countries of the world to cater for an increase in birth rate and the demand for a more skilled work force (Connell, 1980). From the growth of knowledge during the first half of the twentieth century, a demand also developed for a major restructuring of the curricula of the schools, not only in the more developed Western countries, but also in those countries that had gained their independence in Asia, Africa and the Middle East and that had been affected by conflict (Connell, 1980). This reforming of the curricula of the schools had barely begun when the launching of Sputnik, a product of the technological advances in the Soviet Union, challenged the Western world and more especially the United States to immediate action. In the past, the planning of education in the schools had been guided by tradition and affiliation with a major empire as well as the advances in the fields of psychology and human development, and subsequently in the 1940s by a growing interest in sociologically based studies. A group of scholars meeting under the auspices of Unesco, in Hamburg at the Unesco Institute for Education recognized how little was known about the processes of education across the world beyond the sporadic studies undertaken in individual countries. Unesco was the organization that had been established after the Second World War to integrate the educational, scientific and cultural activities of countries across the world in the building of peace and unity among countries and cultures (Husén, 1967).

It was in this climate that an International Association for the

Evaluation of Educational Achievement was born alongside what emerged as a massive expansion of research into the problems of education. The initial goal of the new International Association that became known as IEA was a study of the process of educating youth in a changing world. The assessment, evaluation and investigation of the learning that resulted from teaching in schools were seen to take place in the natural settings that existed in the different countries across the world. This approach was challenged by scholars in the fields of (a) Comparative Education (Getzels, Lipham, & Campbell, 1968), who recognized the educational and cultural traditions of the different countries, (b) Experimental Psychology, who contended that sound knowledge about education could only be built from small experimental studies, (c) Sociology, who viewed education as ideologically based, and (d) Economics, who viewed with concern the rising costs of the provision of education and meeting the demand for an increasingly skilled workforce. However, the initial vision of the small group of scholars in Hamburg has flourished and IEA has gradually undertaken the task of providing empirical evidence for both planning and evaluating the work of transforming education across the developed and developing countries of the world. These scholars have been greatly assisted in their research endeavours by advances in computing, information and communication technology and travel by air.

This article tells the story of the first 33 years of this work during the 50 years of life of IEA. It is a story of a struggle with limited finances, of challenges from both outside and inside the field of education, of the intrusion of political issues, and a lack of appropriate procedures for the statistical analysis of the growing bodies of data that were being collected by IEA and kindred organizations. The relationships between IEA and the resolution of the last named aspect of the struggle are dealt with in an article in the second volume. In this article, the studies undertaken by IEA in its first 30 years are discussed in chronological order and an example of a major policy recommendation arising from each project is given.

### *A Personal Comment*

In the early 1960s, I was a member of the staff of the Australian Council for Educational Research with responsibility for the introduction of new curricula, learning programs and educational materials, particularly in mathematics and the sciences. I received from England a newsletter that included a brief account of a study in 12 countries that had examined the achievement outcomes of education and the article made a closing reference to a future study in the field of mathematics. Twelve months later, I was asked to work on this study of mathematics achievement at the 13 year-old and terminal secondary school levels in the government schools in five of the six Australian states. Testing would take place in Australia in August 1964, six months after testing had been undertaken in the northern hemisphere. Copies of the tests would be sent out from England after use in English schools to reduce costs and to avoid the printing of tests in Australia. However, the optical mark scored answer sheets and the punched cards would need to be sent to Chicago within six weeks after testing for processing and the results tabulated for a meeting in Chicago in January 1965. This was the first empirical research study that I had been asked to carry out during my brief time at ACER. The nature of the study, the use of optical mark scoring answer sheets and punch cards, the computer analysis of data and participation in a meeting over a two-week period in Chicago, coupled with attendance at an American Educational Research Association Conference, were to transform my understanding of educational research from both international and national perspectives.

### **The Founding Meeting (1958) and the Pilot Project (1959-61)**

At the founding meeting in Hamburg in 1958, William D. (Bill) Wall, the Director of the National Foundation for Educational Research in England and Wales (NFER), Torsten Husén from the University of Stockholm, Georg Rasch from Denmark, and Willard Olson from the

University of Michigan in the United States among others, were present. A decision was made to meet again in Hamburg in June 1959 and then a week later at Eltham Palace in England to examine the possibility of conducting an exploratory international study of educational achievement and to consider what might be gained from such an initial investigation. The important meeting was in England and Benjamin S. (Ben) Bloom, from the University of Chicago, Arthur W. (Wells) Foshay, Robert L. (Bob) Thorndike and A. Harry Passow, from Teachers College, Columbia University in New York, Douglas Pidgeon from NFER in England, Gaston Mialaret from France and Fernand Hotyat from Belgium were also present. Wall and Husén would appear to have been the driving forces in the planning of this seminal study that became known as the Pilot Project (1959-61) (Husén, 1967).

The Pilot Project sought to investigate in 12 countries (see Table 1 for the names), the outcomes of educational achievement in reading comprehension, mathematics, science, geography and non-verbal ability. The target population was the last age level at which nearly all of an age group remained at school in the countries involved in the study, aged 13.00 to 13.11 years. The samples tested were not random probability samples, but judgement samples of approximately 1000 students to whom a test of 120 items was given. The index of magnitudes of cross-national differences between subject areas was estimated by the variance between countries as a percentage of the typical variance within a country. The largest value was in mathematics (16%) and the smallest in science (5%) and with the value of 12 per cent for non-verbal ability, that was considered to be a culture free test with the expectation of a very small between country variance. The Pilot Study established that a cross-national investigation was feasible in spite of the problems of translation and administration, since the findings were meaningful. Moreover, the study gave rise to many hypotheses that could be tested in future studies (Foshay, Thorndike, Hotyat, Pidgeon & Walker, 1962).

## **The Cross-National Study of Mathematics (1964)**

As the study of mathematics progressed from the initial decision to embark on an assessment of achievement in mathematics in June 1960, there were significant changes in the composition of the working group of scholars from the universities and research institutes in the countries involved. In general, the members of the group were the leaders of the national research institutes or senior professors in the fields of education in the major universities in each affiliated member country. While the study remained located at the Unesco Institute for Education in Hamburg, it continued as a separate identity with its own financial arrangements. Torsten Husén replaced Bill Wall as Chairman and Technical Director of IEA. An application for financial assistance was submitted to the United States Office of Education and was supported. This led to the full-time appointment of a research worker to coordinate activities and guide the Mathematics Study. Neville Postlethwaite, who was a member of the staff of the National Foundation for Educational Research in London, was appointed to this position of Project Coordinator and subsequently became Director of the study. He not only had the remarkable ability of speaking and writing in several of the foreign languages involved, but he also had excellent administrative skills and the capacity to work extremely hard and quickly, as well as to inspire his colleagues to work hard to achieve seemingly impossible goals. Torsten Husén was an excellent chairman of the Association who had contacts across Europe and the United States and the capacity to integrate the efforts of all taking part and to develop compromises when differences of opinion arose.

The research scholars in the Association gave their time voluntarily and willingly. The costs within each participating country were covered by an organization within that country, and the international costs were provided by the grant from the US Office of Education. There was little prior knowledge or experience to follow except that obtained in the Pilot Project and the experiences gained within some



of the countries that had conducted surveys of educational achievement within their own country. Previous comparative studies in education had been largely concerned with the collation and discussion of descriptive information relating to the education systems operating in different countries and a completely new approach was required.

In this study of achievement in mathematics it was planned to use quantitative and psychometric techniques as well as random probability sampling at several levels of secondary education. The study set out to ascertain the educational objectives of the school systems in the different countries and their relationships to the outcomes. Testing was conducted at three levels that were chosen because they represented terminal points in the teaching and learning of specific groups of students in the countries involved. The lowest level groups of students were 13 years of age, with conjoined samples of students in their eighth year of formal education, who were approaching the end of their period of obligatory schooling. Samples from an intermediate population of 15 and 16 year-old students were also tested in some countries to assess the mathematical competence reached by students before commencing training for vocational employment. At a third level, tests were administered to students who intended to graduate from secondary schools and who would be eligible for university studies or other forms of higher education. They formed two groups: those who studied mathematics, and those who did not.

From the outset of the study a conceptual framework was developed and various committees working within the framework produced a series of hypotheses. These hypotheses were grouped under the following four main headings: (a) the effects of school organization on mathematics achievement, (b) curriculum and instruction, (c) support for education, and (d) the relationships of social, economic and geographic factors to mathematical achievement (Husén, 1967). The study was planned and conducted at a time when the so-called 'new mathematics' was being introduced and it was possible to examine the changes made to the mathematics curricula of the different countries.

Mathematics had been chosen as the first field for inquiry into the school curriculum, because it had been accepted that there was more in common in the field of mathematics across the countries involved than in any other subject. However, major differences between the Francophone countries and other Western nations emerged, based on their different mathematical traditions. Bulletins were developed and issued that spelt out the design, hypotheses, the methods of analysis and dummy tables for the presentation of results in the reports to follow. The teams of scholars, who had formulated the hypotheses for the four areas given above, were assigned to the tasks of writing the associated chapters for the reports that gave the main cross-national findings of the study. These books were published in 1967 by John Wiley in New York and Almqvist and Wiksell in Stockholm with Torsten Husén as editor and with different sections being written by different members of the team of scholars working on the study (Husén, 1967).

The design of probability samples at the different levels of schooling chosen was a challenging task. The Association was fortunate to obtain the services of Gilbert Peaker, who held the position of one of Her Majesty's Inspectors (HMI) in England and Wales and who had lectured in statistics at Cambridge University, as well as being responsible for the sampling in the Plowden National Survey that was in progress in Britain. The samples were necessarily stratified by region, with schools as the primary sampling unit and with a specified number of students selected randomly from within each school. The estimation of sampling errors, not only required random selection, but also allowance made for stratification and a cluster sample design. This was achieved by the forming of four subsamples that were forwarded in four separate batches to Chicago. These batches were used to estimate the magnitudes of the standard errors for the different statistics that were estimated.

Across the 12 countries (see Table 1 for names) over 133,000 students from 5,400 schools were tested, and the questionnaires were completed by 13,000 teachers and 5,400 head teachers (Keeves, 1968).

Consequently, the task of data processing was enormous. Four questionnaires had been constructed: a student questionnaire, a teacher questionnaire, a school questionnaire and a national information questionnaire. Seventy pieces of background information on each student, 40 on each teacher and 30 on each school were recorded on punch cards for processing by computer at the University of Chicago (Keeves, 1968). This work of data processing was carried out by Richard M. (Dick) Wolf and his colleagues who were graduate students in the Measurement, Evaluation and Statistical Analysis Program under Ben Bloom in the Department of Education at the University of Chicago. These students were also assigned the task of interpreting the computer printout provided for each of the scholars in the team who were involved in writing the reports for publication. They were an excellent group and this is reflected in their subsequent distinguished careers.

The construction of the final tests in Mathematics was undertaken by Douglas Pidgeon and Bob Thorndike and the development of the attitude and descriptive scales was carried out by Dick Wolf and his student colleagues. Participating countries were asked to supply test and questionnaire items, but it was only through access to the collections of test items from the University of Chicago and from Educational Testing Service at Princeton, and statements for the attitude and descriptive scales from the University of Chicago, that suitable instruments were constructed.

The tests and other instruments were strongly criticised by both mathematicians and psychologists, who opposed the whole enterprise of assessing the outcomes of education by employing objective tests and attitude scales that could be processed by computer through the use of optical mark scored answer sheets. These were novel in the early 1960s. In addition, Arnold Andersen, a sociologist at the University of Chicago, developed an occupational scale for use across the 12 countries involved in the study to examine the effects of socio-economic status on student performance. There was also some opposition to the

use of such a scale in an educational research study. There is little doubt that the instruments employed drew on the extensive experience of Bob Thorndike at Teachers College, and Ben Bloom at the University of Chicago. Likewise, the Taxonomies of Educational Objectives (Bloom, 1956; Krathwohl, Bloom & Masia, 1964), that were still in the stages of development and publication, influenced the design and structure of the instruments which were prepared for use in the Mathematics Study.

Originally it was planned that analysis of covariance procedures would be used to test the hypotheses that were advanced in the bulletins of the study. These analytical procedures were developed at the University of Chicago to enable this to be done, and were later published, in 1974, by Jeremy Finn (Finn, 1974; Finn & Mattsson, 1974) who was a member of Dick Wolf's team that undertook both the data processing and the analysis. However, a stepwise regression analysis program was released in the early 1960s in the Bio-Medical Program series and after an extended discussion, it was acknowledged that the use of this program was the most appropriate way of analysing the data, although it did not make provision for the nested nature of the data.

The findings of the Mathematics Study reported in 1967 (Husén, 1967) showed that much could be learned from an empirical approach to cross-national research, particularly with respect to the issue of the provision of comprehensive schooling in contrast with selective schooling. This issue was thoroughly examined and this led to two further issues, namely (a) socio-economic status in selective systems, and (b) achievement and retentivity, together with the development of a model of the effects of selection. These ideas were developed further in a separate volume by Neville Postlethwaite (1967) that introduced the idea of 'yield' and addressed the issues of ability grouping and grade repetition. These issues continue to be important more than 40 years later, and the ideas of 'retentivity' and 'yield' are considered further in the second volume of this history of IEA.

When the reports of the Mathematics Study were released in 1967, an emphasis was unfortunately placed by the media on the rank ordering of the achievement in mathematics of the different countries that was contrary to the original purposes of the study. This release of the findings took place shortly after the reports of the Equality of Educational Opportunity Study (Coleman et al. 1966) and the Plowden National Survey (Peaker, 1967) and there was considerable interest in the methodological and analytical issues raised. The results of the Mathematics Study required careful and thorough examination and a conference was held at Lake Mohonk in the United States (Super, 1967) to discuss these findings, that resulted in an extended model of 'Input, Process, Output and Utilization in a Social and Cultural Setting.' This model served to guide the next major undertaking by IEA. The model was complex and causal in structure and raised issues that were clearly important. However, the statistical procedures that were required to examine the relationships in this extended model were not yet available. The nature of the model developed at the Lake Mohonk Conference in 1967 is considered in greater detail in the second volume. Nevertheless, plans for the undertaking of further work by IEA teams of scholars were soon in progress following the holding of a planning meeting in Hamburg in November 1966, and more detailed work on the development of the design of the next round of studies, again in Hamburg, took place in December 1967. Table 1 records the countries that participated in the Pilot Project in 1960, the Mathematics Study in 1964 and the two stages of the Six Subject Study in 1970-1971.

#### **Major Implication for the Planning of Mathematics Education**

All school systems suffered to some extent from lack of equity between different groups of students. This could be rectified by appropriate policies related to gender, social status and rurality leading to higher retention rates.

**Table 1.** Participation in IEA Studies from 1960 to 1971

Study/ Country	Pilot Study	Mathematics (FIMS)	Six-Subject Study (1970-71)					
			Science	Reading	Literature	French	English	Civics
Age of Testing		13	14	14	14	14	14	14
Australia		*	*					
Belgium(Fl)a		*	*	*	*		*	
Belgium(Fr) b	*	*	*	*	*		*	
Chile			*	*	*	T	T	
England	*	*	*	*	*	*		
FR Germany c	*	*	*	*			*	*
Finland	*	*	*	*	*		*	*
France	*	T	T					
Hungary			*	*			*	
India			*	*				
Iran			*	*	*			*
Ireland								*
Israel	*	*		*			*	*
Italy			*	*	*		*	*
Japan		*	*					
Netherlands		*	*	*		*	*	*
New Zealand			*	*	*	*		*
Poland	*							
Rumania						*		
Scotland	*	*	*	*		*		
Sweden	*	*	*	*	*	*		T
Switzerland	*							
Thailand			*				*	
United States	*	*	*	*	*	*		*
Yugoslavia	*							
Number	12	12	19	16	10	8	10	10

a (Fl) Flemish b, (Fr) French c, FR Federal Republic, T Terminal secondary school level only.

FIMS refers to the First International Mathematics Study conducted in 1964.

## **The Six Subject Study (1970-71)**

The Six Subject Study was planned as a very ambitious project in two stages to cover the major areas of the school curriculum. The study was designed to include the 10 year-old students in primary schools, the 14 year-old students in middle secondary schools and students at the terminal secondary school level. The first stage assessed student performance in Science (Comber & Keeves, 1973), Reading Comprehension (Thorndike, 1973), and Literature (Purves, 1973), conducted in 1970, and the second stage assessed performance in French as a Foreign Language (Carroll, 1975), English as a Foreign Language (Lewis & Massad, 1975) and Civic Education (Oppenheim & Torney, 1974; Torney, Oppenheim & Farnen, 1976) and was conducted in 1971. It was recognized that each subject field would take four to five years to plan and four years for the analysis of data and writing a report that would be published as a separate volume. In addition, there would be a system case-study report, prepared under the leadership of Harry Passow, with help from Harold Noah, Max Eckstein and John Mallea (Passow, Noah, Eckstein & Mallea, 1976), together with an overview or summary report, written by David Walker (1976), that covered all six subject areas. In addition, there would be a technical report prepared by Gilbert Peaker (1975) who was responsible for the sampling, data processing and data analysis.

The achievement tests and the attitude scales were planned to be Rasch scaled under the direction of Bruce Choppin, with the watchful guidance of Bob Thorndike and Dick Wolf and with bridging items to link together the three different levels of testing. This work proved to be too ambitious, because appropriate programs had not been developed to carry out Rasch scaling with large bodies of data and the data processors reverted to the use of classical test and Likert scale scoring procedures. The explanatory analyses were undertaken in Stockholm, and canonical analysis and path analysis were considered as possible approaches to the multivariate analysis of the data. These plans gave

rise to work by Jöreskog (Jöreskog & Sorbom, 1978) using maximum likelihood procedures and Hermann Wold (1982) using partial least squares with latent variables procedures, but these procedures were not used in the main analyses of the data. However, these procedures were employed by Ingrid Munck (1979) using LISREL, and Richard (Dick) Noonan (1976) using LVPLS in their secondary analyses for their publications in the IEA Monograph Series. The secondary analysis of data became a notable feature of the Six Subject Study and was strongly supported by the Spencer Foundation with Jeremy Finn and Dick Noonan both working on Spencer Fellowships. This program involved five groups of five Spencer Fellows who each worked in Stockholm for a 12 month period.

Only the French as a Foreign Language Study reported by John B. (Jack) Carroll (1975), who employed canonical analysis, and the Science Study, reported by L.C. (Sam) Comber and John Keeves (1973), who presented a simple example of path analysis, used the originally proposed analytical procedures in the analyses of the Six Subject Study data in the main reports. In 1971, Gilbert Peaker (1975) published a monograph from a follow-up study of the students tested in the Plowden National Survey in England that was the first known use of path analysis in educational research. The group working in Stockholm acknowledged the problems arising from the multilevel nature of the data and discussed with J. Coleman appropriate procedures for the analyses at the student and school levels which were used under Gilbert Peaker's direction in the analyses for the six subject areas. However, no analytical strategy could be devised at this time to combine these data for analysis at two or three levels. These and other statistical issues, that were faced and tackled in IEA studies are considered in greater detail in an article in the second volume on the relationship of the IEA studies to educational research and educational policy and practice.

With the expansion of the IEA program to over 20 countries, four of which were referred to as developing countries (Chile, India, Iran and



Thailand) in contrast with the Western or developed countries, and the limited participation of countries from Eastern Europe it was considered desirable to expand IEA activities in a way that would spread knowledge of the approach employed for the assessment of educational outcomes both by objective and performance tests. In the foreign languages, for example, performance tests of Speaking, Listening, Reading and Writing were assessed. In the Reading Comprehension Study (Thorndike, 1973), Reading, Speed of Reading and Word Knowledge (Verbal Ability) were assessed, and in the Science Study, Practical Work (Doran & Tamir, 1992) and an Understanding of Science (Comber & Keeves, 1973) were investigated, in response to the criticism of the limited nature of the objective tests assessed in the Mathematics Study.

Consequently, in mid-1971 a workshop was held in Gränna in Sweden on 'Curriculum Development and Evaluation' for teams of participants from 23 countries. The workshop was conducted by Ben Bloom with the help of Ralph Tyler and IEA colleagues using examples drawn from the current IEA studies and recorded in the recently published book *Handbook of Formative and Summative Evaluation of Educational Student Learning* (Bloom, Hastings & Madaus, 1971). The workshop also drew on the *Taxonomies of Educational Objectives* that were published in 1956 on the Cognitive Domain (Bloom, 1956) and in 1964 on the *Affective Domain* (Krathwohl, Bloom & Masia, 1964). In addition, Ralph Tyler's seminal work that was published in 1949 on *Basic Principles of Curriculum and Instruction* was republished for the Gränna Workshop (Tyler, 1949). IEA thus contributed to the advancement of not only the strategies of assessment employed across the world, but also the development of curricula in the countries involved in the IEA program of research. Furthermore, the conduct of a study that spread across six curriculum areas tended to reduce the view that the IEA program was an educational Olympiad with countries ranked in their order of performance on the IEA tests. Nevertheless, this simplistic view of the IEA studies subsequently led critics not only to

attack the tests employed, but also the errors of measurement involved, rather than recognize that IEA was primarily concerned with the conduct of research and the building of a greater body of knowledge about education across the world, and not with the ranking of countries with respect to their levels of achievement.

### **Relocation of IEA to Stockholm and Legal Incorporation of IEA (1969)**

The expansion of the IEA program of research and evaluation demanded that greater office space was available for visiting scholars, and for meetings as well as access to more powerful computer facilities. Torsten Husén saw the possibilities within the University of Stockholm and the Wenner Gren Center to relocate the IEA Headquarters to Sweden within the University from the Unesco Institute in Hamburg. In 1969, this relocation of offices was smoothly accomplished. Subsequently, offices within the nearby university buildings became available and in mid 1971 IEA moved into university buildings, with access to the residential facilities of the Wenner Gren Center. For seven years until 1978, while Torsten Husén was Chairman of IEA this provided an international centre for IEA as well as opportunities for graduate students to take courses leading to higher degrees, with a substantial number completing doctorates that also involved the presentation and defence in a public disputation of a published work that was linked to the IEA research program. For a further 12 years the IEA offices remained nominally in Stockholm within the Institute of International Education under the guidance of Ingemar Fagerlind and Torsten Husén, although the Chairmen, namely Neville Postlethwaite (1978-1986) and Alan Purves (1987-1989), worked in other locations.

The initiation of the Six Subject Study led to the need to establish IEA as a legal entity within a country that permitted subsequent movement of the seat of operation to another country that might from time

to time change. This was possible in Belgium where Gilbert de Landsheere was a highly esteemed member of the staff of the University of Liege. In addition to receiving grants from the United States Office of Education, substantial support was received from the Ford Foundation, the Leverhulme Foundation as well as the Spencer Foundation and subsequently from a foundation established by Robert Maxwell who had established Pergamon Press in England.

IEA gradually acquired considerable standing across the countries of the world for its high level of scholarship and research, in the fields of Comparative Education, Curriculum Development, and Educational Policy Making. Robert Maxwell saw the need to publish widely on educational issues and sought Torsten Husén's help and advice about the publishing of an *International Encyclopedia of Education* with supplementary volumes issued at successive intervals. Torsten Husén clearly saw the demand for such a source of information on the findings of research and scholarly writing and agreed to accept the responsibility of editorship with the help of Neville Postlethwaite. Together with the wide-spread network of colleagues in IEA it was possible to embark on this challenging enterprise. In return, Pergamon Press agreed to publish the reports of IEA studies and to establish two periodicals, *Studies in Educational Evaluation* and *the International Journal of Educational Research* that would be edited from within the IEA network which had been built up by Torsten Husén and Neville Postlethwaite. These publications can be said to have transformed the scholarly study of education across both the developed and developing countries of the world, together with an ongoing flow of publications written personally by Torsten Husén and published largely through Almqvist and Wiksell in Sweden. Other European publishing houses and in particular, Elsevier have continued this work in the years that have followed, keeping abreast of developments with new forms of information and communication technology and with a steady flow of handbooks and periodicals. In addition to the two editions of the *International Encyclopedia of Education* as well as the elec-

tronic version of the *Complete Encyclopedia of Education*, eight handbooks associated with the first edition of the Encyclopedia and eight associated with the second edition were published. While these volumes are not actually IEA publications, they were informally linked to IEA. Furthermore, at least 13 books have been published by Pergamon Press that have been related to IEA projects and on the methods of IEA research. Moreover, 11 books reporting on IEA research were published jointly during a 30 year period by Almqvist and Wiksell from Stockholm and John Wiley from New York, while an annotated bibliography from 1960 to 1990 of IEA publications listed over 600 entries (Degenhart, 1990).

### **The Science Study (1966-1973)**

The team that was initially assembled to plan the Science Study in 1966 involved people working in the field of science teaching, namely L.C. (Sam) Comber, a biologist and HMI from England, William (Bill) Ritchie, a physicist and HMI from Scotland, Richard Whitfield, a chemist from Cambridge, England, William Hered and Leopold (Leo) Klopfer, from Harvard and Chicago in the United States, Karl Hecht, from the IPN in Kiel in Germany, Shigeki Sakakibara from NIER in Japan and myself, John Keeves, Federal Secretary of the Australian Science Teachers Association, and an educational research worker who had taught science in schools in Australia and England. All were actively involved in the many different aspects of the reform movements in schools to redevelop the school science curricula in our own countries. We sought to map the curricula with respect to content and scientific skills that were being introduced across the countries under survey, and we sought to assess an understanding of the processes of science that were held by students in schools. The tests and the attitude and descriptive scales that were constructed were perhaps too wordy, but the new ideas that were being introduced into science education required careful wording. Sam Comber worked tirelessly to

perfect the instruments. We were perhaps too ambitious in what we attempted to do and we left many important aspects of the originally planned study incomplete and the data collected not fully analysed.

However, while the main study was in progress, I worked on a complementary study (Keeves, 1972) to examine the educational environment of the home, the classroom and the school, and the peer group with an Australian sample of 13 year- old students in order to identify the factors that influenced the learning of science. This Australian study was longitudinal in nature across a year of school learning in science and was published in 1972 by Almqvist and Wiksell in Stockholm from the IEA headquarters.

While the international study was essentially cross-sectional in nature, we recognized that with the three age levels under survey, it was possible to assess the amount of learning taking place between age levels in each of the countries. Figure 1 records the findings taken from the published report on the increase in level of performance in science from the 10 year-old age level to the terminal secondary school stage (Comber & Keeves, 1973). From Figure 1, four important findings can be clearly seen. First, the performance in science of the four developing countries, Chile, India, Iran and Thailand were approximately a student standard deviation lower in the level of the achieved learning of science than in the 14 developed countries. Second, countries differed greatly in the level of achievement at the 10 year-old stage with the greatest difference between Japan and Germany being of the order of one student standard deviation. Third, countries differed greatly in the gain between the 10 year-old (Population I) and the 14 year-old (Population II) age levels in achievement with two kindred countries, Hungary and Finland showing the greatest difference. Fourth, the level of achievement at the terminal secondary school (Population IV) level was related to the proportion of the age group remaining at school to the terminal stage (note in Figure 1: the countries are listed in terms of retention rates at the last year of schooling). Some of the issues associated with factors influencing performance and change in

performance are considered in a later section of this article. The factors at the initial stage appeared to be related to ideas advanced in Jack Carroll's (1963) model of school learning associated with time and opportunity to learn, and this idea was developed by Ben Bloom (Lietz, 1996) in the dissemination of the findings from this study. The multivariate regression analyses that were undertaken at the student and school levels gave rise to many important findings related to the teaching of students in classrooms.

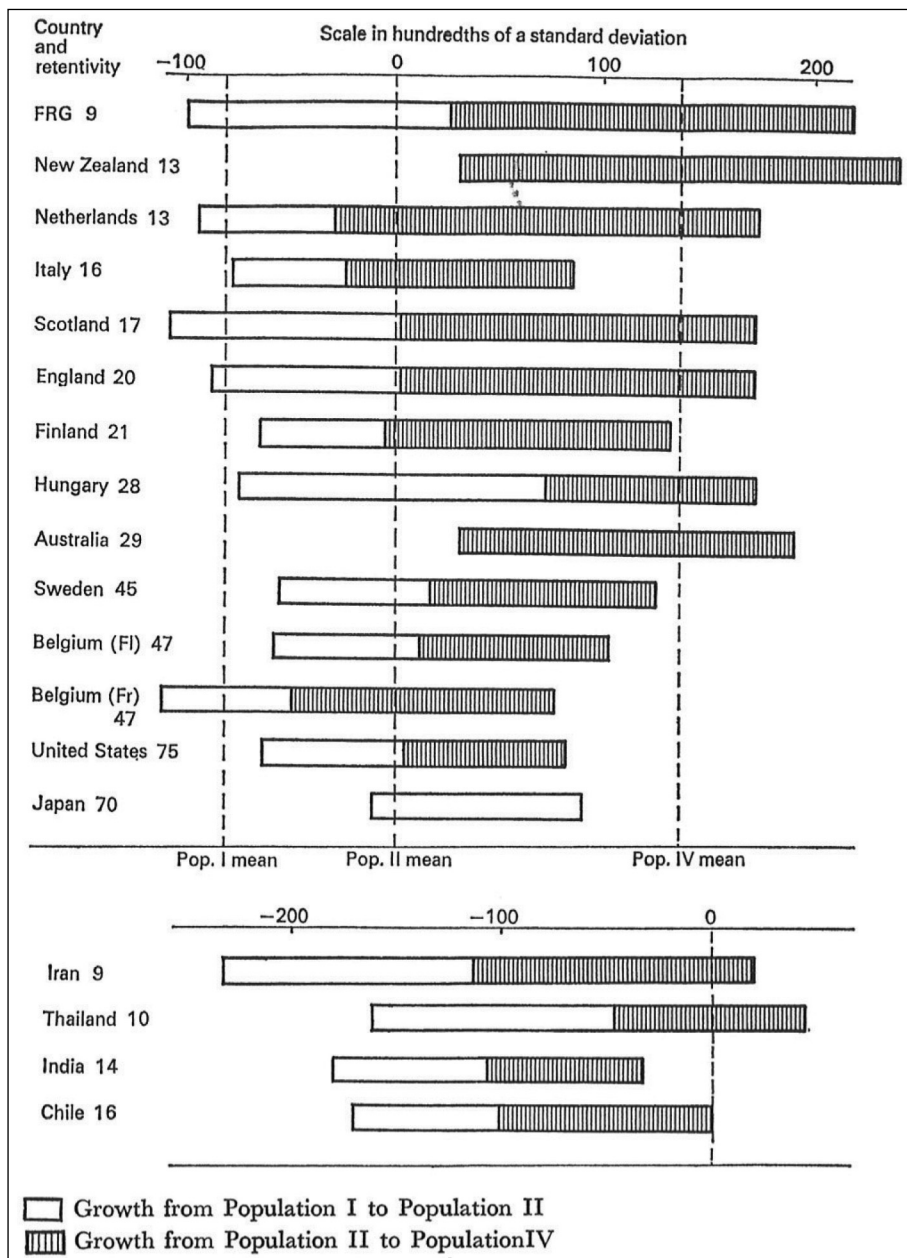
The IEA studies were greatly limited by the procedures available for the statistical analysis of data to identifying separately the student and school level factors influencing science learning. An opportunity was sought 14 years later to undertake a trend study in order to examine change in the teaching and learning of science over a critical period in science education across the 18 countries involved in this initial study of science achievement. Of particular interest from the findings of the First IEA Science Study was the low performance of students in developing countries (see Figure 1). Lawrence (Larry) Saha, a visiting Fellow, and Ingemar Fagerlind, worked at the International Institute of Education in Stockholm on the preparation of a study *Education and National Development: A Comparative Perspective*, that was published in two editions by Pergamon Press (Fagerlind & Saha, 1989).

#### **Major Implication for the Planning of Science Education**

Competent and committed science teachers are needed for the successful teaching of science in secondary schools.

#### **The Reading Comprehension Study (1970-71)**

This study was directed and reported by Bob Thorndike (1973) who had a long-standing interest in the field and the issues addressed. Tests of Reading Comprehension, Speed of Reading, and Word Knowledge, which was considered to be a measure of verbal ability, were developed and translated from English into 11 languages. Variables were examined as possible explanatory factors of reading achievement in



**Figure 1.** Increase in level of performance in science from the ten-year-old level to the terminal secondary school stage (taken from Comber&Keeves, 1973, p. 171)

six areas: (a) the out-of-school environment, including home environment and language of the home, (b) availability of reading materials at school, (c) instructional practices, (d) interests and attitudes, particularly with respect to reading, (e) acquired study and reading habits, and (f) presence of physical impairments, and those of sight, hearing or speech. Three populations were tested across 15 countries.

There was a lack of explanatory power of school and classroom based variables in accounting for the differences between students and schools within a country. The stronger effects of home-related factors suggested that either schools were equal in their provision of opportunities for the development of reading abilities at the upper primary and lower secondary school levels, or that the focus of mother-tongue instruction beyond the early grades was not on the furthering of reading comprehension skills. Apart from the differences in level of reading achievement between the more developed countries and the developing countries at the 10 year-old and 14 year-old levels, the differences between countries in reading achievement were small. However, at the terminal secondary school level differences between these countries in levels of reading achievement could be largely accounted for by differences in retentivity (Walker, 1967). Bob Thorndike, like his father before him, argued that 'reading beyond the fifth grade level involved largely reasoning'.

#### **Major Implication for Education in Reading Comprehension**

Schools need to conduct programs for the development of reading comprehension skills at the upper primary and secondary school levels.

### **The Harvard Conference (1973)**

The publications from the first stage of the Six Subject Study in Science by Comber and Keeves (1973), in Reading Comprehension by Thorndike (1973) and in Literature by Purves (1973) were presented and discussed at a Conference on Educational Policy and International



Assessment at Harvard University in November 1973 and reported by Purves and Levine (1975). This conference could be said to have re-oriented the IEA program of research towards the making of educational policy.

### **The Reformation of IEA, Tokyo (1978)**

For three years after the completion of the Six Subject Study, the Spencer Fellows worked at the Institute of International Education within the University of Stockholm under the guidance of Torsten Husén. Neville Postlethwaite moved from Stockholm to the Unesco International Institute for Educational Planning in Paris, and in 1976 he was called to a professorship of comparative education at the University of Hamburg. This appointment provided him with the opportunity to take on responsibilities for the running of IEA, and at the General Assembly meeting in Tokyo in 1978, he became Chairman of the Association. Torsten Husén relinquished the chairmanship after 17 years, during which two major research studies had been conducted with 20 member countries and IEA was beginning to have a significant impact on policy making in education throughout the world. Torsten Husén travelled widely, wrote prolifically, beginning each day with several hours of writing, which he was able to do in at least four languages, from a base in Sweden that was politically independent of both East and West, as well as North and South affiliations.

Rebuilding IEA in the late 1970s was no easy task when Neville Postlethwaite took over as Chairman. There was no money in the bank and money was difficult to raise from the United States, unless the study was administered from within that country. A Second International Mathematics Study (SIMS) was planned, that was led by Kenneth (Ken) Travers from the United States and David Robitaille from the University of British Columbia in Canada. A chapter on the SIMS project is included in this volume. A Classroom Environment Study was initiated under the leadership of Nathaniel (Nate) Gage at

Stanford University that was later relocated to the Ontario Institute for Studies in Education under the guidance of Bernard Shapiro. However, the data analysis was largely undertaken at the University of Hamburg in order to employ the analytical procedure of path analysis with latent variables, where Norbert Sellin, a graduate student, extended the program developed by Hermann Wold in Sweden. In addition, a Second IEA Science Study was advanced from the Australian Council for Educational Research and Malcolm Rosier became coordinator of the project. However, within universities and research institutes there was sufficient interest in research that examined the processes of education, in spite of the fact that financial support was very limited, for these research activities to proceed.

### **The Second IEA Science Study (1979-1989)**

The Second IEA Science Study was planned in the late 1970s as a trend study to be conducted in 1983-84, approximately 14 years after the first study in 1970. The purpose of this study was primarily to investigate the changes that had taken place in the teaching of science as a consequence of the major changes that had occurred around the world during the 1960s and early 1970s in science education. The team of scholars who worked on the project were drawn largely from the research institutes that had been involved in the conduct of the first study. They had had the experience of undertaking a cross-national research study in science and understood the issues associated with both policy and practice in the organization of the teaching and learning of science from a national perspective. As a consequence, planning meetings of national research coordinators (NRCs) were held annually in the centres involved in the study, namely NIER in Tokyo (2 meetings), CIER in Rome (2 meetings), IEP in Budapest (1 meeting), ACER in Canberra and Melbourne (1 meeting), Manilla (1 meeting), Teachers College, Columbia University in New York (1 meeting) and OISE in Toronto (1 meeting). National Centres were financially responsible not only for

the conduct of the study in their own countries, but also for travel to the meetings of the NRCs for the planning of the study. Those national centres that hosted meetings covered the accommodation costs of visiting NRCs. Limited financial support came from a Japanese Foundation, the Sasakawa Foundation, the World Bank, and Australian sources to cover international costs. The emphasis in the reporting of the study was on the national reports, and each country involved was encouraged to prepare its own report and to write and support the publication of articles on the study. Although the writing of the major reports was delayed, 120 articles on the study were published prior to mid-1990 before the main reports were completed. There were, however, three major developments that occurred in the statistical analysis and measurement that needed to be taken into consideration. First, Rasch scaling was evolving that would permit the measurement on a common scale both across occasions and the three age levels at which assessment took place. Second, procedures for the multivariate analysis of data using path analysis with latent variable were becoming available to replace the regression analysis procedures that had been employed in the earlier IEA studies. Third, the problems associated with the multilevel nature of the data were being tackled using both least squares and maximum likelihood procedures.

There was clearly a need for cross-national comparisons of science achievement but it proved impossible to raise sufficient financial support within Australia. Consequently, it became necessary to relocate the international work to Europe and in particular to Stockholm, where support was obtained from the Bank of Sweden Tercentenary Foundation and Svenska Handelsbanken, and in Hamburg, where the National Science Foundation in the United States was able to provide a grant through Teachers College, Columbia University, in New York.

The tests and attitude scales had been developed with sufficient bridge and anchor items to link across the three year levels of schooling, and across the two occasions of testing for Rasch scaling to be employed. However, when the work was undertaken in Hamburg by Andreas

Schleicher, a student of Neville Postlethwaite, anomalies arose because the only computer program available was not capable of doing what was required and had to be redeveloped. As a consequence, most of the scaling and analysis had to be redone. Likewise, for the multivariate analyses at the student and school levels, the program initially available, LVPLS, while satisfactory, did not record adequately the indirect effects that had a mediating influence on the outcome and Norbert Sellin developed a program PLSPATH to carry out this work. With respect to the multilevel problems, initial analyses were carried out with the help of Ray Adams in Australia that involved ordinary least squares analysis operating at the school and classroom levels (Keeves & Larkin, 1968). Subsequently, with data from Hong Kong, exploratory work was done by Kwong (K. C.) Cheung and Norbert Sellin using both MLWin and HLM/2L in Hamburg to test these procedures, but the demands on computer operating time were too heavy for extensive use of these techniques with the large bodies of data involved in the Second IEA Science Study (Cheung, 1990).

Several international reports were ultimately published. The first brief report (IEA, 1988) examined the achievement of students in each country as well as their relative performance across countries. The second report (Rosier & Keeves, 1991) described the curriculum and conditions of learning in the countries involved, as well as providing an account of the changes that had occurred during recent decades in the teaching and learning of science. The third report (Postlethwaite & Wiley, 1992) was a more detailed statement obtained from the analysis of data on the achievement of the outcomes of science teaching in the countries involved. The fourth book (Keeves, 1992a) examined the changes in performance and the conditions of the teaching and learning of science in the ten countries that had participated in both the First and Second IEA Science Studies. The final report was titled *Learning Science in a Changing World* (Keeves, 1992b), and it summarized the key findings from this major study as well as the changes occurring over time and across the three levels of schooling. The test-

ing of practical skills was considered in a separate report that added a new dimension to the IEA program of research. This important part of the Second IEA Science Study was undertaken in six countries and was reported in a special issue of *Studies in Educational Evaluation* with Rod Doran and Pinchas Tamir as editors (Doran & Tamir, 1992).

### **Major Implication for Planning Science Education**

Because of the increasing emphasis on the learning of science and the continuing growth of scientific knowledge, student achievement and the teaching and learning of science at all grade levels should be monitored by education systems in order to support the development of science education across the world.

### **The IEA Reading Literacy Study (1990-91)**

This study was planned through the initiative of Neville Postlethwaite at the University of Hamburg as a follow-up study that was undertaken 20 years earlier on Reading Comprehension. Funding was so limited that the staff working on the study could only be issued with monthly contracts. Moreover, a special not-for profit organisation, the "IEA-Hamburg e.V." had to be created in order to employ staff legally without losing to the university much of the scarce funding available.

Warwick Elley guided the conduct of the study, while Dieter Kotte and Petra Lietz coordinated activities in Hamburg, Andreas Schleicher undertook the data processing under the supervision of Albert (Al) Beaton from Educational Testing Service in the United States, and Ken Ross provided advice on the sampling. The purposes of this study were to: (a) describe the achievement levels in reading literacy in education systems; (b) describe voluntary reading activities of 9 year-old and 14 year-old students; (c) identify policies and practices that related to reading achievement and voluntary reading; (d) produce valid international tests for investigating reading development; and (e) provide national baseline data for monitoring literacy levels over time.

For this study, reading literacy referred to the linguistic and cognitive processing of those written language forms required by society and valued by the individual. Testing took place at the 9 and 14 year-old age levels. Performance in Reading Literacy was calibrated on Rasch measurement scales for use in literacy testing programs in 32 countries, thus laying the foundations for future studies of reading literacy at these age levels.

The single most critical factor in the development of literacy was the educative or literate environment of the students' homes (Lundberg & Linnakyla, 1993). However, students whose home language differed from the language of instruction had lower literacy performance in all countries at both age levels (Elley, 1992, 1993). Girls achieved at higher levels than boys in all countries at the 9 year-old level and in most countries at the 14 year-old level (Elley, 1992). Students in classes where teachers emphasized reading for comprehension and encouraged more reading did better (Postlethwaite & Ross, 1992). Subsequently, the data for the Reading Comprehension Study in 1970-71 were equated to the data for the Reading Literacy Study in 1990-91 and levels in performance over time were examined across those eight countries that had participated on both occasions at the 14-year-old level (Lietz, 1996).

#### **Major Implication for Planning the Development of Reading Literacy**

In order to increase student reading performance, voluntary out-of-school reading should be fostered among students, particularly during the primary school years. Schools should have classroom or school libraries and teachers should emphasize reading for comprehension.

#### **Issues Arising from the First 30 Years of Operation**

IEA was heavily dependent on sources in the United States, both governmental and non-governmental, for the funding required to conduct

the international aspects of the program of research and evaluation. Where an operating centre for a project was located in the United States, it was possible to channel money through that centre. However, where the operating centre was outside the United States difficulties arose since IEA was governed by law from Belgium, and from 1958 to 1990 the administration headquarters were located first in Hamburg and then in Stockholm. The International Institute for Education in the University of Stockholm received substantial financial support from Swedish sources and was guided directly by Torsten Husén, initially as Chairman of IEA. When Neville Postlethwaite took over as Chairman of IEA the Headquarters remained in Stockholm and when he retired from the position of Chairman, Alan Purves, the new Chairman, was located in the United States, but with limited contact with Stockholm. The issue arose as to whether or not the Headquarters should move to the United States. However, a decision was made for Tjeerd Plomp to assume the Chairmanship of IEA and he was not only able to receive financial support from sources in the Netherlands, but was also able to set IEA on a firmer financial footing by the introduction of membership fees and to continue as a powerful force in educational research world-wide.

### *A Challenge from the United States*

From the mid 1960s when it was recognized that the Mathematics Study would release a significant report, the Educational Testing Service at Princeton in the United States, saw the work undertaken by IEA as not involving the monitoring of student achievement across countries but as a research and evaluation enterprise. Consequently, it set up a counter association referred to as IAEP (The International Association for Educational Performance) that conducted cross-national assessment and monitoring programs. This became a direct challenge to IEA and Torsten Husén argued for IEA building a liaison with the National Assessment Program of Educational Progress (NAEP) at Educational Testing Service. This led to Al Beaton estab-

lishing ties with IEA and subsequently becoming involved in the Reading Literacy Study led by Neville Postlethwaite from Hamburg. Subsequently, Al Beaton left Educational Testing Service to work with IEA in Boston. This move to some extent broke the links between IEA and ETS, and in due course IAEP ceased involvement in cross-national monitoring studies.

### ***The World Bank***

The World Bank was impressed with the work of IEA not only in developed countries but also in developing countries and sought opportunities to support studies being undertaken in developing countries. In 1981, a major development was proposed at the General Assembly meeting in Canberra, Australia to build a branch program of IEA's work in developing countries. This move was rejected from within the United States, by scholars who questioned whether this was the appropriate support that the developing countries required. Subsequently, in 1989, William (Bill Loxley) from the Asian Development Bank staff became director, of IEA's research and development programs.

### ***Training Activities of IEA***

The Gränna Workshop in Sweden in 1971 can be said to have initiated IEA's work in the training of research workers in curriculum evaluation and assessment. As a consequence, IEA began to recognize the importance of its work for other aspects of empirical research, such as providing training in sampling and the estimation of sampling errors as well as data processing and statistical analysis. There is little doubt that graduate students who became involved in IEA studies in Chicago, Stockholm, Hamburg and Melbourne have gone on to make major contributions not only in statistical analysis (eg. Finn, Noonan, Sellin), evaluation (eg. Wolf), conducting monitoring studies (eg. Schleicher with PISA, Ross with SACMEQ) and educational measurement (eg. Choppin with Rasch scaling, and Adams with QUEST and



CONQUEST). The Second IEA Science Study saw IEA's role in training as being a major activity. In 1992, an *IEA Technical Handbook* was prepared in Stockholm from both new articles and material previously published by Pergamon Press and was distributed by IEA from The Hague on behalf of the Technical Advisory Committee (Keeves, 1992c).

### *Data Processing Center*

Neville Postlethwaite, working in Hamburg from 1976 until retirement in 1995, saw the importance of establishing a centre that focussed on data processing of the highest quality. Individuals and teams working under his guidance had developed data processing procedures that had maintained the highest possible standards. From these roots, a large centre was subsequently built in Hamburg that specialized in such work and which became the IEA Data Processing and Research Center, led by Dirk Hastedt and Heiko Sibberns.

### *The Eastern Bloc Countries and China*

In the Pilot Study (1959-61) two Eastern European countries namely, Poland and Yugoslavia were involved, but were unable to continue participation in subsequent IEA projects. However, Romania became involved in the French as a Foreign Language Study (1971). Moreover, Hungary became very actively involved in the First IEA Science Study (1970-71) and led by Zoltan Bathory contributed in substantial ways to the Second IEA Science Study (1979-1989). With the opening up of Eastern Europe after 1989, ten countries became participating members of IEA from Eastern Europe prior to mid-1994. In addition, from the International Institute for Education in Stockholm support was provided for Zhau Shangwu from China to undertake a replication of the Second IEA Science Study that has subsequently greatly influenced China's involvement in IEA (Shangwu, 1993).

### *Spin Off Programs*

While IEA has continued to thrive since the end of the period of

approximately a third of a century under survey in this chapter it is interesting to note that participants in IEA studies have been recruited to develop similar programs that served different purposes. Andreas Schleicher with outstanding organizational skills has been responsible for setting up and maintaining several programs undertaken by the Organization for Economic Cooperation and Development (OECD) in Paris including the Programme for International Student Assessment (PISA). In addition, Ken Ross has worked, from the Institute for International Educational Planning (IIEP) in Paris, on the organization of the Southern and Eastern African Consortium for Monitoring Educational Quality (SACMEQ) that operated in Sub-Saharan Africa. The World Bank continued to be active with support for Neville Postlethwaite in the conduct of separate studies in Eastern Asia. However, the focus of these so-called 'Spin-Off Programs' has largely been on monitoring performance rather than the quest for knowledge about the processes of education. Nevertheless, there is an immense body of data that has been collected over the years that awaits further analysis with the analytical procedures that are now available, although many problems remain (Deping, Osanje & Jiong, 2009; Rutkowski et al, 2010).

## Conclusion

The program of research and evaluation conducted by IEA over a 30 year period responded not only to the need for greater accountability in education systems in countries across the world, but also in transforming the field of Comparative Education. Through the research studies reported in IEA publications and the *International Encyclopaedia of Education*, its handbooks and the associated journals, IEA has gradually transformed empirical research in education into a scholarly enterprise across the countries of the developed and developing world. The founders of IEA saw a need for an empirically oriented, cooperative research program in education that investigated problems

that were common across the national systems of education. They viewed the world as a global laboratory where different national policies and practices produced different educational outcomes.

The IEA studies initially advanced and tested a large number of hypotheses, which were subsequently built into models that demanded testing by multilevel and multivariate procedures of statistical analysis. The questions addressed were increasingly more complex and simple answers could not generally be obtained. However, the financial support for IEA research studies came from those administrators and policy makers who sought advice for the development of policy and the guiding of practice. Under these circumstances they turned to IEA and its kindred research programs to provide that advice. There was clearly a need for the research workers engaged in IEA studies in each country to provide recommendations that served the needs of each country involved. Consequently, a short monograph titled *The World of School Learning: Selected Key Findings from 35 Years of IEA Research* sought to summarize the results of IEA studies and to provide recommendations for educational planning and policy making (Keeves, 1995).

Perhaps the greatest shortcoming of the research undertaken over this period of a third of a century was the failure to take into consideration the cultural differences that operated between and within countries and national systems of education. Moreover, the initial vision of building a body of theoretical knowledge about educational processes that had been tested against evidence from the real world remained as a major challenge for scholars in Comparative Education at a time when there was growing interest and concern for globalization. However, the findings obtained from IEA research studies and those of kindred programs have been so great that the task of addressing the problems raised initially in Hamburg over 50 years ago, remains an immense, challenging and perhaps endless quest.

## References

- Bloom, B. S. (Ed.). (1956). *Taxonomy of Educational Objectives. The classification of educational goals. Handbook 1. Cognitive domain*. New York: McKay.
- Bloom, B. S., Hastings, J. T. & Madaus, G. F. (1971). *Handbook of Formative and Summative Evaluation of Student Learning*. New York: McGraw Hill.
- Carroll, J. B. (1963). A model of school learning. *Teachers College Record* 64, 723-733.
- Carroll, J. B. (1975). *The Teaching of French as a Foreign Language in Eight Countries*. Stockholm: Almqvist and Wiksell, and New York: John Wiley.
- Cheung, K. C. et al. (1990). The Analysis of Multilevel Data in Educational Research: Studies of Problems and their Solutions. *International Journal of Educational Research*, 12, 217-317.
- Coleman, J. S. et al. (1966). *Equality of Educational Opportunity. Volumes I and II*. Washington: US Government Printing Office.
- Comber, L. C. & Keeves, J. P. (1973). *Science Education in Nineteen Countries*. Stockholm: Almqvist and Wiksell.
- Connell, W. F. (1980). *A History of Education in the Twentieth Century World*. New York: Teachers College Press.
- Degenhart, R. E. (1990). *Thirty Years of International Research. An annotated bibliography of IEA publications (1960-1990)*. The Hague: IEA.
- Deping L., Osanje, A., & Jiong, Y. (2009). On the Estimation of Hierarchical Latent Regression Models for Large-Scale Assessments. *Journal of Educational and Behavioral Statistics*, 34, 433-463.
- Doran, R. L., & Tamir, P. (Eds.). (1992). Practical Skills Testing: *Studies in Educational Evaluation*, 18, 263-406.
- Elley, W. B. (1992). *How in the World Do Students Read?* The Hague: IEA.
- Elley, W. B. (1993). *The IEA Study of Reading Achievement and Instruction in Thirty-Two School Systems*. Oxford: Pergamon Press.

- Fagerlind, I., & Saha, L.J. (1989). *Education and National Development: A Comparative Perspective* (2nd ed.). Oxford: Pergamon Press.
- Finn, J. D. (1974) *A General Model for Multivariate Analysis*. New York: McGraw Hill.
- Finn, J. D., & Mattsson, I. (1974). *Multivariate Analysis in Educational Research: Some applications*. Stockholm: Institute of the Study of International Problems in Education, University of Stockholm.
- Foshay, A. W., Thorndike, R. L., Hotyat, F., Pidgeon, D. A. & Walker, D. A. (1962). *Educational Achievements of Thirteen-Year-Olds in Twelve Countries*. Hamburg: Unesco Institute for Education.
- Getzels, J., Lipham, J., & Campbell, R. (1968). *Educational Administration as a Social Process*. New York: Harper and Row.
- Husén, T. (Ed.) (1967). *International Study of Achievement in Mathematics*. Stockholm: Almqvist and Wiksell.
- Jöreskog, K. G. & Sorbom, D. (1978). *LISREL IV Analysis of Linear Structural Relationships by the Method of Maximum Likelihood*. Chicago: National Educational Resources.
- Keeves, J. P. (1968). *Variation in Mathematics Education in Australia*. Hawthorn, Vic.: ACER.
- Keeves, J. P. (1972). *Educational Environment and Student Achievement*. Stockholm: Almqvist and Wiksell.
- Keeves, J. P. (Ed.). (1992a). *The IEA Study of Science III: Changes in Science Education and Achievement: 1970 to 1984*. Oxford: Pergamon Press.
- Keeves, J. P. (1992b). *Learning Science in a Changing World. Cross-national Studies of Science Achievement: 1970 to 1984*. The Hague: IEA.
- Keeves, J. P. (Ed.). (1992c). *The IEA Technical Handbook*. The Hague: IEA.
- Keeves, J. P. (1995). *The World of School Learning: Selected key findings from 35 years of IEA research*. The Hague: IEA.
- Keeves, J. P. & Larkin, A. (1986). The context of academic motivation. *International Journal of Educational Research*, 10, 205-213.

- Krathwohl, D. R., Bloom, B. S., & Masia, B. B. (1964). *Taxonomy of Educational Objectives, The classification of educational goals. Handbook 2. Affective domain*. New York: McKay.
- Lewis, E. G. & Massad, C. E. (1975). *The Teaching of English as a Foreign Language in Ten Countries*. Stockholm: Almqvist and Wiksell.
- Lietz, P. (1996). *Changes in Reading Comprehension across Cultures and over Time*. Münster/New York: Waxmann.
- Lundberg, I., & Linnakyla, P. (1993). *Teaching Reading around the World*. The Hague: IEA.
- Munck, I. M. E. (1979). *Model Building in Comparative Education*. IEA Monograph Studies No. 1. Stockholm: Almqvist and Wiksell.
- Noonan, R. D. (1976). *School Resources, Social Class and Student Achievement: A Comparative Study of School Resource Allocation and Social Distribution of Mathematics Achievement in Ten Countries*. Stockholm: Almqvist and Wiksell.
- Oppenheim, A. N. & Torney, J. V. (1974). *The Measurement of Children's Civic Attitudes in Different Nations*. IEA Monograph Studies No. 2. Stockholm: Almqvist and Wiksell.
- Passow, A. H., Noah, H. J., Eckstein, M. A. & Mallea, J. R. (1976). *The National Case Study: An Empirical Comparative Study of Twenty-One Educational Systems*. Stockholm: Almqvist and Wiksell, and New York: John Wiley.
- Peaker, G. F. (1967). The Regression Analyses of the National Survey. In Central Advisory Council for Education (England) *Children and their Primary Schools*. (Plowden Report). London: HMSO. Vol II Appendix IV. pp. 179-221).
- Peaker, G. F. (1975). *An Empirical Study of Education in Twenty-One Countries: A Technical Report*. Stockholm: Almqvist and Wiksell, and New York: John Wiley.
- Postlethwaite, T. N. (1967). *School Organization and Student Achievement*. Stockholm: Almqvist and Wiksell.

- Postlethwaite, T. N., & Ross, K. N. (1992). *Effective Schools in Reading: Implications for Educational Planners*. The Hague: IEA
- Postlethwaite, T. N. & Wiley, D. E. (1992). *The IEA Study of Science II: Science Achievement in Twenty-Three Countries*. Oxford: Pergamon Press.
- Purves, A. C. (1973). *Literature Education in Ten Countries*. Stockholm: Almqvist and Wiksell.
- Purves, A. C., & Levine, D. C. (1975). *Educational Policy and International Assessment: Implications of the IEA surveys of achievement*. Berkeley, CA: McCutchan.
- Rosier, M. J. & Keeves, J. P. (Eds.). (1991). *The IEA Study of Science I: Science Education and Curricula in Twenty-Three Countries*. Oxford: Pergamon Press.
- Rutkowski, L. et al. (2010). International Large Scale Assessment Data: Issues in secondary analysis and reporting. *Educational Researcher*, 39(2), 142-157.
- Shangwu, Z. (1993). *Chinese Science Education: A Comparative Study of Achievement in Secondary Schools related to Student Home and School Factors*. Stockholm: IIE, University of Stockholm.
- Super, D. E. (Ed.) (1967). *Towards a Cross-national Model of Educational Achievement in a National Economy*. New York: Teachers College, Columbia University.
- The International Association for the Evaluation of Educational Achievement (1988). *Science Achievement in Seventeen Countries. A preliminary report*. Oxford: Pergamon Press.
- Thorndike, R. L. (1973). *Reading Comprehension Education in Fifteen Countries: An Empirical Study*. Stockholm: Almqvist and Wiksell, and New York: John Wiley.
- Torney, J. V., Oppenheim, A. N. & Farnen, R. N. (1976). *Civic Education in Ten Countries*. Stockholm: Almqvist and Wiksell.
- Tyler, R. (1949). *Basic Principles of Curriculum and Instruction*. Chicago: University of Chicago Press. (republished, 1971).

- 
- Walker, D. A. (1967). An attempt to construct a model on the effects of selection. In T. Husén, *International Study of Achievement in Mathematics*, Vol 2, (pp. 135-9). Stockholm: Almqvist and Wiksell.
- Walker, D. A. (1976). *The IEA Six-Subject Survey: An Empirical Study of Education in Twenty-One Countries*. Stockholm: Almqvist and Wiksell, and New York: John Wiley.
- Wold, H. (1982). Soft Modeling: The basic design and some extensions. In K. G. Jöreskog, & H. Wold (Eds.), *Systems under Indirect Observation Part II*. Amsterdam: North Holland.





## CHAPTER 2

# IEA from 1988-1994: An Organization in Transition

*Tjeerd Plomp*

### Introduction

It was under the leadership of Torsten Husen, IEA Chair from 1959 until 1978, that the reputation of IEA as *the association* for the international comparative assessment of educational achievement was established. His successor, Neville Postlethwaite, Chair from 1978 till 1986, was working in a downward economic climate, in which it was not easy to acquire funds for the kind of studies that IEA would undertake. However, under Postlethwaite's leadership, the international study coordinators (and their centers), the chairpersons of international steering committees of the various IEA studies and several IEA member countries all made a concerted effort to acquire funding for IEA's international overhead and its studies. I can state with impunity that Postlethwaite was a 'strong captain on a stormy sea'.

Towards the end of the 1980s, the IEA was running a number of studies simultaneously, although each was at a different stage of the process. While the Second International Mathematics Study (SIMS) was completed in 1986, both the Second International Science Study (SISS) and the Written Composition Study were still in progress, formally ending in 1988; the Pre-Primary Project (PPP) and the Computers in Education

---

This chapter on IEA is written from the perspective of the various positions I held in the last part of the 20th century. From 1987-1989, I was GA representative for The Netherlands and IEA Chair from January 1, 1990, until September 1999. In addition, I chaired the International Steering Committees of the Computers in Education study (COMPED) and the Third International Mathematics and Science Study (TIMSS).

(COMPED) Study were in full swing; and the Reading Literacy Study was still in the organizational stage, with the first National Research Coordinators meeting scheduled for November 1988.

At that time, due to IEA's 'loose' organization and the lack of a solid financial base, the General Assembly (GA) did not function as a strong governing body and there was no clear coordination in the scheduling and design of studies. Although the Standing Committee (SC) was supportive and the Technical Advisory Committee provided methodological input, basically each study had to do its own fundraising and operated quite independently, each with an international coordinator and national centres participating (see also the chapter by John Keeves, and that of Jack Schwille).

It was under the chairmanship of Alan Purves, my direct predecessor from 1986-1990, that IEA became a 'learning organization'. And it was during the same period that IEA examined and analyzed its organization, leading to the GA decision at the 1988 meeting (in Frascati, Italy) that it was crucial that IEA becomes more stable, both organizationally and financially. That year I was acting as the GA representative for The Netherlands and I became actively involved in these discussions. So from 1990 onwards, as Chair of the IEA, I - together with the Standing Committee (SC) - was responsible for the process of the transition that IEA went through in the 1990s.

In this chapter I will reconstruct, from my personal perspective and background in IEA, some aspects and elements of the first years of this transition process. First I will describe how I perceived the changes in the IEA, and I will also attempt to summarize the discussions within the organization that ultimately led to the decision to change IEA's structure and organization. I will then describe the establishment of the IEA Headquarters - subsequently called the Secretariat<sup>1</sup> - in The Hague (which later moved to Amsterdam) and also how the new

---

1. The terms Secretariat and Headquarters will be used interchangeably.

membership system developed. Then I will specify more precisely the ways in which elements of IEA's transition occurred, as these were crucial to the eventual successful transformation. These included the development of a program of studies (with special attention to how TIMS changed into TIMSS), the creation of the IEA Data Processing Center (DPC) and the development of a publication policy. The chapter will finish with some final remarks and acknowledgements.

### **A Changing Environment Leading to the Transformation of IEA - the 1988 GA Meeting in Frascati (Italy)**

In the late 1980s many national governments showed an increased interest in monitoring educational quality as part of their policy. A number of these (e.g., England and Wales, Hungary, Sweden, the USA, Canada/ British Columbia and The Netherlands) had already introduced regular (sample-based) assessments of student achievement in primary and/or secondary education. At that time we saw - (in addition to the IEA studies) - an increasing involvement of nations in international educational indicator projects. For example, the OECD study of educational indicators took off - resulting in the well-known publication "Education at a Glance" - whilst a number of countries participated in the International Assessment of Educational Progress (IAEP) initiated by the USA and conducted by the Educational Testing Services (ETS) in Princeton, New Jersey. Although those projects had their merits, it was clear to those of us involved in the IEA at that time, that national governments or educational testing centers were not equipped to conduct studies that tested what students from many countries had learned ('attained' curriculum) and process data about schools and the 'implemented' curriculum. The type of studies that the IEA had conducted - such as the SIMS (Second International Mathematics Study) and SISS (Second International Science Study), but also assessments in other areas - was needed to provide data on these more complex sets of variables. In other words, in the late 1980s,

the IEA— given its expertise and experience – was in the unique position to provide highly relevant research data not only to the research community, but also and especially to national governments and international organizations. However, a necessary condition for making the most of this opportunity was lacking-- having an organizationally and financially strong and smoothly running organization.

To understand the IEA context in the late 1980s, one has to realize that the General Assembly was comprised of members who represented either a national research centre or a university that was involved in one of the IEA studies. These member institutes were not under any obligation to represent ‘their’ country in the IEA or, conversely, to represent the IEA in their country. This type of membership implied that the national IEA centres were in no position to pay contributions to the international coordination of studies, nor to commit themselves to paying large fees to establish and maintain an IEA infrastructure. At this time, membership dues were US\$450 per year, to cover the travel expenses of the Chair and part of the salary of a secretary at the Secretariat in Sweden; at this time, too, the IEA Chair did this job in his free time. It was also a time – as Clare Burstall (England) and Inger Marklund (Sweden) remarked in their memo to the 1998 GA meeting in Frascati -- that the IEA had acquired “a reputation for glacial slowness in delivering results”, that national reports more often than not had to be published without international data, that studies started without guaranteed international funding, that too many of the studies had been ad hoc researcher initiatives, and that there was no guarantee that any study would be completed.

In light of this situation, the Standing Committee at that time, comprising Chair Alan Purves (USA) and the members Zoltan Bathory (Hungary), Gilbert De Landsheere (Belgium), Robert Garden (New Zealand), Inger Marklund (Sweden) and Moegiadi (Indonesia), had the vision to understand this situation and to place the need for change prominently on the agenda of the 1988 General Assembly (GA) meeting in Frascati, Italy.

In his 'Opening Remarks' to the 1988 GA meeting, Alan Purves introduced the question about the future of the IEA in the following way:

To pose the issue in the starkest way, it is up to the GA this week to decide whether to disband IEA as soon as the current studies come to their conclusion. If the decision is to continue IEA and proceed with future studies, there must be a commitment from all members to support the international research activities, both financially and by fulfilling the various obligations of membership in insuring prompt response to requests for information, care in carrying out the research agenda, and participation in all international activities. It has been the experience of the International Coordinating Centers as well as of the International Secretariat<sup>2</sup> that such full commitment comes from about half of the national centers; the others are passive members.

At the same meeting, Burstall and Marklund (1988) noted a considerable number of points in favour of continuing IEA, including, for example, that IEA was internationally recognized in its field, provided countries and systems with a broader perspective on educational policy and practice than that offered by purely national studies, had fostered the establishment of an international research community, had increasingly become salient to international policy-makers, and had provided important training in survey methodology. They recommended that the IEA should consider cycles of repeat surveys in domains such as mathematics, science, mother tongue and foreign language teaching, as the IEA not only had demonstrated its strength in these areas, but that it was also likely that these subjects would continue to be important. They also recommended that data collected in IEA studies should be related to policy and to the strong move to educational accountability, which also constituted the basis for national assessment programmes. They also expected that there would be an international appetite for data from cycles of repeat surveys – today

---

2. At that time located at the University of Stockholm, Sweden.

we would add, that this contributes to the continuous monitoring of quality of education.

Thus, on the one hand, the 'environment' of the IEA became more and more suitable for IEA type of international comparative achievement studies, whilst on the other hand the IEA leadership realized more and more that a reorganization would be needed to meet the challenges of 'serving the world's expectations'.

After fruitful discussions on IEA's mission and future –based on the Chair's position paper and the Burstall and Marklund memo– not only did the the 1988 GA meeting in Frascati approve drafting of a new mission statement, but the GA and the Standing Committee decided to reorganize IEA's infrastructure by (i) founding an International Secretariat to coordinate all IEA activities, and (ii) introducing a fee system to finance the costs of the international secretariat as well as the international coordination centers of IEA studies. It was noted that a fee system should reflect the relative wealth of a country, and that contributions in kind (such as hosting meetings) should be an option for members unable to to pay a fee.

It is noteworthy that the discussions at the GA meeting also addressed the issue of a 'centralized versus decentralized' structure, resulting in a decision that the permanent secretariat should be located near a strong IEA centre and be able to provide centralized data processing capabilities for the IEA studies. The Standing Committee also decided to create sub-committees for both aspects of the reorganization; these were expected to report to the Standing Committee at its February 1989 meeting (in Stockholm, Sweden).

One may conclude that the 1988 General Assembly was a turning point in IEA's history. At its meeting of February 1989 (in Stockholm, Sweden), the Standing Committee elaborated on the decisions of the 1988 GA, by stating that a single IEA Headquarters, operating ultimately under the leadership of the IEA chair, would provide the following:

1. coordination among all projects in terms of design and scheduling (implying that all International Study Centers would work through the IEA Headquarters)
2. handling all international data processing and preparation (e.g., manuals, codebooks, etc.)
3. coordination of all fundraising
4. handling all publications and public relations
5. maintenance of all archives.

The SC decided at the same meeting – as it was considering candidates to host the new Headquarters – to invite the Netherlands Foundation for Educational Research (SVO) in The Hague to offer a proposal for housing the IEA Headquarters, whilst it invited me to be available as IEA Chairman beginning January 1, 1990 – to be approved by the IEA General Assembly meeting of September 1989 in Seoul, Republic of Korea.

### **Establishing the IEA Secretariat**

It was not only the Netherlands Foundation for Educational Research (SVO) that showed commitment to the new policy initiatives of the IEA; the Netherlands Ministry of Education and Sciences (MES) -- commissioner of the IEA studies in the country -- had also shown increasing interest in national and international assessment and indicator studies. Therefore, in line with this interest, the Netherlands MES supported the idea of housing the IEA Secretariat in The Netherlands.

As it was important to give the Secretariat time to establish itself and prove its value to the IEA, the SVO and the MES were willing to commit themselves for a first period of four years, with an option for extension. Although the SVO was closed in 1997, the commitment of the Netherlands MES became a long-term one, and continues to this day.



The proposal for the IEA Secretariat presented in March 1989 by the SVO took as starting points the five tasks listed by the SC (see above). Without going into detail about the secretarial, financial and publication/PR tasks, it is – in retrospect – interesting to see how *data management* was elaborated to include:

- supporting international coordination centers (ICCs) of studies in designing data management and processing for that particular study in line with IEA standards and guidelines
- planning and executing between-study comparisons
- setting up instrument collection to serve new studies
- data archiving
- organizing training sessions for new ICCs and NRCs

In other words, this proposal provided a first statement of mission and purpose for the IEA Data Processing Center that was established in 1993.

The proposal suggested that for the first four years, the IEA Secretariat would be hosted by the SVO (Netherlands Foundation for Educational Research) and would consist of the chairperson, and full-time positions for a senior executive officer (the IEA director), a data manager and an administrator,, whilst the SVO would take care of IEA's financial administration. The first annual budget for the IEA Secretariat added up to Dfl 600,000 (≈ €272,700) of which – after some negotiations - the Dutch Ministry of Education and Sciences (MES) would cover Dfl 130,000 (21.7%), SVO Dfl 100,000 (16.7%; partly donations in kind), whilst my university (i.e., University of Twente, Enschede, The Netherlands) was willing to invoice only a small fraction of my salary costs. This arrangement would give the IEA four years to develop its membership system and to generate the remainder of the budget through its fee system (to be established) and/or the generosity of donor agencies.

The IEA accepted the proposal of the SVO and the Dutch MSE; thus,

the IEA Secretariat in The Hague became operational in 1990 with the appointment of Bill Loxley as director and José Ammerlaan as secretary. In 1992, Wim Hayes was appointed as office manager, and was succeeded by Barbara Malak in 1995 who became Manager of Membership Relations. Andreas Schleicher (International coordinator of the Reading Literacy Study) started on January 1, 1993, as the staff member responsible for data management. Bill Loxley left the IEA in 1994 and was succeeded – after an interim period – in May 1997 by the present Executive Director Hans Wagemaker, who had already been involved in the IEA for many years in various capabilities, including GA representative, member of the Standing Committee and member of the International Steering Committee for the Reading Literacy Study.

In 1996 the IEA Secretariat moved to its present premises in Amsterdam and today has a staff of eight people, whilst the Data Processing and Research Center in Hamburg is formally also part of the IEA Secretariat.

The growth of the IEA since then can also be illustrated by comparing the balance sheets of 1990 and 2008 (per December 31): assets per December 31, 1990: US\$166,022; and for 2008: US\$ 8,138,917 (source: IEA Secretariat).

### **The Development of the Membership and Fee System**

When I was appointed in 1989 as Chair, the IEA Guidebook of that year listed 33 education system as members of the IEA, representing 31 countries (as Belgium was represented with two education systems, Flanders and Wallonia, as was Canada with the two provinces of British Columbia and Ontario). When I resigned the chairmanship in 1999, the IEA had 55 members with only Belgium represented with two education systems. This growth suggests the organization's successful transitional in the 1990s. I would like to argue-- although I cannot provide empirical data --that this growth may be ascribed to the

organizational and financial restructuring of the IEA in combination with the development of an attractive program of studies of good quality. In the following sections, I will elaborate on some of these points, but in this section I want to report on some of the struggles we (i.e., jointly the Standing Committee and the Secretariat) had in the early 1990s in implementing the restructuring of the IEA.

Before summarizing how these issues appeared at consecutive GA meetings, it is important to point to the fact that successful organizational change requires reforms in the 'culture' of the organization as well as a change in the attitude and orientation of the individual GA members and their centres.

In retrospect, it appears that, although a majority of the GA representatives had no difficulties in accepting the organizational and structural changes decided upon at the 1988 and 1989 GA meetings, a significant number of the GA members had problems meeting the new membership commitments (such as raising funds for the IEA Secretariat and for the international coordination of the studies). This was not surprising, as for many years the IEA was an organisation conducting studies initiated and driven by the interest of researchers, whilst the members were not 'suffering' from raising the funds needed for the international overhead (as stated, the membership fee in 1988 was US\$450). But as of 1988, IEA aimed to become an organization with a program of studies generated on the basis of the mission of the IEA, a related policy plan and the commitment of the members to assure the international costs of both the Secretariat and the studies.

I am convinced that when in 1988 a substantial number of GA members voted wholeheartedly for the 'future' of the IEA they did not realize what this would mean with regard to their position as GA representatives. This can be illustrated in the discussions and motions on the financial base of the IEA at the 1989 GA meeting in Seoul (Rep of Korea). After intense discussions, the GA proposed a number of motions (see minutes of the GA), including: (i) that individual educa-

tion systems contribute to IEA activities within Headquarters and projects (22 voted for, 1 abstention); (ii) that for a period of three years as of September 1989 the basic dues for Headquarters activities be US\$10,000 annually per (national) centre, with payment to be negotiated between IEA Headquarters and each centre (15 for, 8 abstentions); and (iii) effective from September 1990, for the five-year active life of any ongoing project, each participating centre contribute to international costs (of the ICC) according to the number of participating centres and the yearly project budget, amounting to not more than US\$10,000 annually with payment to be negotiated between IEA Headquarters and each centre (12 for, 1 against, 10 abstentions). The voting on the first motion illustrates that (almost) every member agreed with the principle of all members contributing to the organization, but the voting on the third motion illustrates a need for proceeding cautiously and sensitively. The Standing Committee was instructed to provide IEA Headquarters with guidelines for negotiations with individual centres and the GA also suggested that Headquarters should explore support from international agencies and private enterprises.<sup>3</sup>

The IEA leadership and the newly established Secretariat actively tried in the early 1990s to lessen the members' financial requirements by exploring partnerships with international institutions like UNESCO, OECD, the World Bank and the European Union to gain their interest, involvement and support for the IEA studies. Both the Chair and the Executive Director conducted many courtesy visits to establish links with such institutions. This can also be illustrated by looking at the international observers at the GA meetings: in 1990 there was only a representative of the Pergamon Press (the 'IEA publisher'), whilst in later years UNESCO, OECD, the World Bank and the European Union

---

3. The Minutes of the 1992 GA meeting (Ascona, Switzerland) mention that two consultants be contracted for fund raising (one from the USA and one from Europe), but I do not remember that this has resulted in meaningful donations to the IEA.

were regularly (if not annually) represented at the GA meetings (see, e.g., the chapter by Marlaine Lockheed for an impression from the World Bank perspective).

As quite a number of countries showed an interest in joining the new IEA studies, the SC proposed to the GA a set of criteria for IEA members so as to ensure that new members would not enter IEA with the wrong expectations. The 1991 GA meeting (Enschede, The Netherlands) discussed the structure of the IEA and agreed that:

to join IEA, a national institute must meet four requirements:

1. be associated with a strong research tradition with linkages to educational researchers and practitioners;
2. have working relationships with national policy makers;
3. have ties to national research funding agencies;
4. represent a total educational system.

It is clear that the SC also wanted this motion adopted to put some pressure on those members unable to meet the contribution expectations.

However, the challenge to have *all* members actively contributing to the IEA continued to be problematic. The SC and the Secretariat generated and tried out many ideas to address this issue. They attempted to involve international institutions in IEA studies, and suggested implementing a fee system that would permit contributions in cash or in kind, having a weighted system such that each member country would be invoiced relative to its economic status, to change the status of non-paying members into that of 'non-voting' member. I suggested in my chairman's report to the 1990 GA meeting (Beijing, PRChina) that members should change their national membership organization to make their country (or education system) a stronger member of the IEA. This was what The Netherlands did, by transferring their IEA membership from a university to the SVO (the National Institute for Educational Research). I referred to a variety of 'models' for arranging the national membership that were existing at that time (besides university research centres representing their country), such as an IEA consortium (Federal Republic of Germany), a national institute for

educational research (Sweden, The Netherlands), directorates of a Ministry of Education responsible for educational research (New Zealand, Norway). Eventually, although difficult for some countries in the beginning, most IEA members were able to find ways to meet the financial obligations of membership, whilst the World Bank -- thanks to the efforts of Executive Director, Hans Wagemaker, and Marlaine Lockheed, the World Bank representative in IEA -- appeared to be willing to support a number of 'economically weak' countries.

Currently three weighing factors are used to determine membership fees: Gross Domestic Product, Purchasing Power Parity per Capita, and Public Spending on Education, based on sources such as World Development Indicators (World Bank) and the CIA World Factbook (UNESCO Statistic). To determine fees for participation in studies, three starting points are used: expected international costs, basic funding that might be obtained from third parties such as the World Bank, European Commission and/or International Development Bank, and the number of educational systems that might participate in a study (source: the IEA Secretariat).

### **Towards a Program of Studies**

At the 1988 GA meeting (Frascati, Italy) one important aspect of the discussion on IEA's future pertained to the need for 'cycles of repeat surveys' in basic subjects ( mathematics, science, mother tongue and foreign languages). It was expected that all countries would express a continuing concern for these subjects when they would have national assessments, and also that there would be an international interest for this kind of data. However, discussion also led to the conclusion that the top priorities in the reorganization should be to establish the IEA Secretariat (or Headquarters) and to plan how this 'central office' could be funded by the membership. In other words, a program of studies would be part of the reorganization, but the IEA first had to reorganize its structure and its operations.

This explains why in the first two years of the transition period most of the attention and energy of the IEA leadership was given to building up IEA's infrastructure. However, at the same time there were items on the GA meeting agenda indicating that the SC was also working on preparing a program of studies.

At the 1989 GA meeting in Seoul, a new mission statement was discussed that resulted in IEA's decision to commit itself "to a cycle of studies of learning in the basic school subjects and to additional studies of particular interest to its members". Important in this mission statement is that space was created for new studies, and later this space was used for studies such as the Teacher Education Study and SITES, the Second Information Technology in Education Study.

Looking at the 'basic subjects', in these years the SC explored what new studies could be initiated - after the Reading Literacy Study that would have data collection in 1990 (northern hemisphere) and 1991 (southern hemisphere) - as steps towards a program of studies. During the 1989 GA meeting (Seoul) a proposal for the Third International Mathematics Study (TIMS) was discussed, and at the 1990 GA (Beijing) the scope of this study was widened to mathematics and science, becoming the well-known Third International Mathematics and Science Study (TIMSS). The 1991 GA meeting (Enschede, The Netherlands) discussed the possibility of a 'second or foreign language' study and passed a motion that the Secretariat should do further preparatory work. As a result, at the 1992 GA meeting (Ascona, Switzerland) it was decided that a foreign/second language study should be designed to survey language proficiency and effective practices among 15-year-olds in the last year of upper secondary school.

The minutes of the 1992 GA meeting reveal that there was discussion about a program of studies, comprising four cycles: (i) Mathematics/Science cycle, (ii) Language cycle, (iii) Civics/Arts cycle, and (iv) Special cycle (with studies like the Computers in Education study). Not surprisingly, the discussions in the GA were intense and ques-

tioned “whether too many large scale studies across many subjects might not be logistically impossible” ( GA minutes, 1992, p.7). A motion was proposed but not voted on by the GA to approve in principle the concept of a new cycle of studies with the understanding that any single study must be approved by the GA in advance (p.7). Instead, the Standing Committee was asked to review the cycle to see how to integrate an IEA cycle with the needs of other international agencies.

At the 1993 GA meeting (Escorial, Spain) the topic was raised as part of the discussion on the future of IEA. The minutes state that:

the Chairperson noted that Reading Literacy [study] was recently completed in 1993, that the Computers in Education study would terminate at the end of 1994, and that the IEA plans to launch its Language in Education Study in 1994 and a Civics in Education Study in 1995. A motion was moved that a new diagram of proposed and current studies would be prepared for the next GA meeting (GA minutes, 1993, p.24).

Such a diagram was indeed presented at the 1994 GA meeting (Yogyakarta, Indonesia) comprising the cycles shown in Figure 1.

It is beyond the scope of my chapter to report on how this cycle of studies developed over the years. I confine myself to a few developments that occurred in the years that followed.

- The *math & science cycle* was implemented very successfully, although the year of data collection of TIMSS had to be moved to 1995 (for the northern hemisphere) and 1996 (for the southern hemisphere), which influenced the schedule for consecutive studies;

- On the *language cycle*: due to lack of funding, the Language of Education Study had to be discontinued; the 1996 GA meeting (Vancouver, Canada) approved preparation of a new Reading Literacy study, which- with data collection in 2001 - became the first ‘Progress in International Reading Literacy’ (PIRLS) study in a five-year-cycle of assessments of trends in reading literacy.



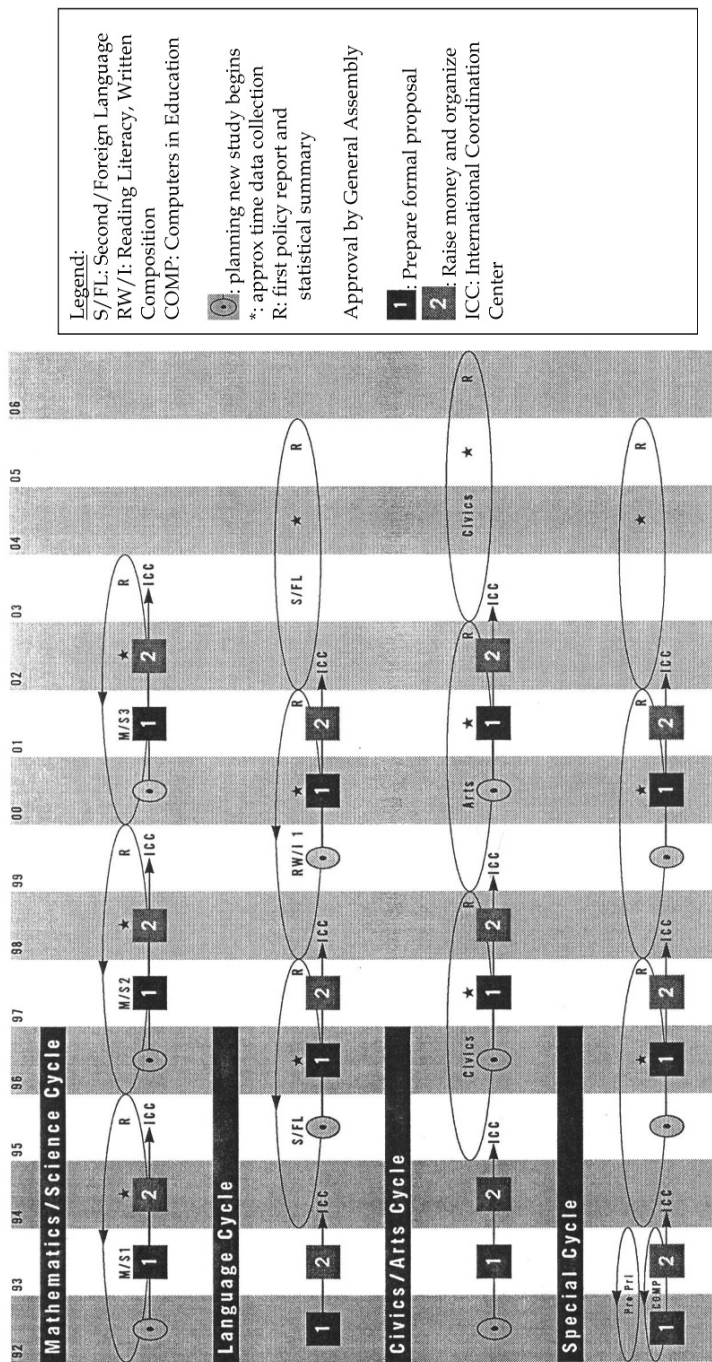


Figure 1. IEA Cycle of Studies (as presented to the IEA General Assembly, 1994 (Yogyakarta, Indonesia))

Studies in both cycles were, and still are, coordinated by the IEA 'TIMSS & PIRLS International Study Center' at Boston College (USA).

- On the *civics/arts cycle*: the IEA did not implement studies on arts education, but a Civics Education Study (CIVED) was initiated with phase 1 conducted in 1996 and 1997, whilst the data were collected in 1999 (and optional in 2000) – see the chapter by Judith Torney and Jack Schwille that discusses the Civics Education studies.

- On the *special cycle*: the 1996 GA meeting (Vancouver) approved a motion that a proposal for the Second Information Technology in Education Study (SITES) be prepared, which resulted in the SITES study comprising various modules with data collection in the years 1998, 2002 and 2006. An example of another study in this special cycle is the Teacher Education and Development Study in Mathematics (TEDS-M).

### **From TIMS to TIMSS**

Above I wrote that during the 1989 GA (Seoul) meeting a proposal for the Third International Mathematics Study (TIMS) was discussed, and at the 1990 GA the scope of this study was widened to mathematics and science, becoming the well-known Third International Mathematics and Science Study (TIMSS). The section that follows will describe the process of the transition from TIMS to TIMSS.

The minutes of the 1989 GA meeting (Seoul) state that “proposals had been invited for the location of the ICC for TIMS. British Columbia responded with a written proposal”. This proposal was written by David Robitaille, the GA representative for Canada-British Columbia, who had been actively involved in the second math study, SIMS. There was a motion for an ad-hoc working group under the chairmanship of Robitaille to be appointed to produce an elaborated proposal for consideration at the 1990 GA meeting. So, we were all heading for a third mathematics study.

After I had taken up the IEA chairmanship in January 1990, my first trip was to the USA, to chair a meeting of the Standing Committee (SC) hosted in Albany (New York) by my predecessor Alan Purves. During that meeting, the SC was informed that the working group for TIMS (only Math) was scheduled to meet in March of that year. The SC also decided in the meeting that Zoltan Bathory (Hungary) would be invited to prepare (with appropriate consultants) an initial proposal for a Third International Science Study (TISS) to be discussed at the 1990 GA meeting in Beijing, China.

A few weeks later, I travelled to Washington D.C., as I had been invited for a courtesy visit to the US Board on International Comparative Studies in Education (BICSE) at the National Research Council. BICSE had become the body that represented the USA in IEA, and in 1990 a member, Gordon Ambach, acted for the first time as the USA GA representative. The creation of BICSE in 1988 showed the USA's increased interest in international comparisons, and their willingness to invest in international assessments. I am referring to the chapters by Jack Schwille, Ken Travers, Gordon Ambach and Larry Suter for more background information about the increasing concern in the USA about the quality of their education from an international point of view, which became manifest in the publication of the report *Nation at Risk* (1983) by the national Commission on Excellence in Education and discussions in the following years.

My visit to BICSE in Washington D.C. was scheduled for February 2, 1990, just after President Bush's 'State of the Union' address on January 31, 1990. In this address, President Bush presented a number of national education goals to be met by the year 2000. One of these goals was that "*by the year 2000 U.S. students must be the first in the world in math and science achievement*". Being an educator from Europe, and not used to this type of 'world championship' language in phrasing national education goals, I remember that I was really surprised, if not perplexed, when I heard about this education goal. Luckily, my American friends explained the metaphorical meaning of it - "Tjeerd,

you cannot expect the President to state that the USA is aiming to become number 2 or number 3". Moreover, when I explained to BICSE, the IEA decision to restructure and set up a program of studies (referring to the plans for a Third International Mathematics Study (TIMS) and a Third International Science Study (TISS), the board expressed its strong support for these studies if IEA were successful in realizing its plans.

A few months later, when I attended the annual conference of the American Education Research Association (AERA) together with the newly appointed IEA director Bill Loxley, we were invited to meet with Emerson Elliott, the Acting Commissioner for Education, and some of his staff, including Jeanne Griffith and Eugene Owen from the National Center for Education Statistics (NCES) of the Department of Education. The NCES was given the task to arrange the "international comparisons of student achievement in mathematics and science to allow us to measure change in the education system during the 1990s. These studies will be used to inform the President of progress towards his goals for American education in the year 2000" (summary quote from a letter by Emerson Elliott to the IEA chair, July 5, 1990). In operational terms, Elliott requested the IEA to merge the planned (separate) studies of mathematics and of science into a combined international comparative achievement study of both domains with data collections in January 1994 and 1998, and with reporting by January 1996 and 2000, respectively. Elliott added to his invitation that - should the IEA accommodate this request - the USA would subsidize a meaningful part of the international coordinating costs for the study.

I still remember how I felt at that time: the dilemma between the 'moral obligation' to consult the SC versus the expectation of our conversation partners for a clear signal from us at this meeting. I also remember that I - being IEA Chair for less than four months -- was considering various points, including: (i) the IEA was already planning studies of mathematics and sciences; (ii) we had no funding yet for the international component of our studies; (iii) that we had here

an IEA member who was willing to subsidize an important part of our plans to establish a program of studies; but also (iv) the concern of some IEA members that one member country could dominate the study because – by paying a major part of the international costs – they might have too much influence over the design of the new study. On the other hand, no other member had expressed such clear confidence in IEA's plans for change combined with the willingness to help it become reality as was proposed. So, I made the on-the-spot decision to consider NCES's offer to be a 'real opportunity' for the IEA and that to decline this offer was not a real option-- as not only would we lose momentum for our own studies, but also the USA would proceed to organize its own international comparative studies of mathematics and sciences, as it had done in the program 'International Assessment of Educational Progress' (IAEP) a few years before. I also was convinced that I – in my capacity as IEA Chair – together with Standing Committee and the proposed International Coordinator and Chair of the International Steering Committee, should be able to make sure that a Third International Mathematics and Science Study (TIMSS) would become a real international comparative assessment, meeting the high standards we had talked about in discussing IEA's future.

After this April meeting with Emerson Elliott and his staff, I informed the members of the IEA Standing Committee, as well as David Robitaille (the international coordinator of TIMS) and the TIMS working group. The SC approved the idea and at the 1990 GA meeting proposed that the IEA embark on a two-subject, two-stage study. David Robitaille (University of British Columbia, Canada) presented the proposal for the combined Third International Mathematics and Science Study (TIMSS) to the 1990 meeting in Beijing, highlighting the design, methodology and other major points for consideration. After ample discussions in small groups and in plenary, the GA unanimously approved a motion to accept the TIMSS design.

The TIMSS ICC (International Coordinating Center) was located at the University of British Columbia, Canada. In 1993, the international

coordination of TIMSS was restructured by appointing Al Beaton (Boston College, formerly ETS) as Study Director in charge of aspects like instrument development, sampling and data analysis. Over the years, Al Beaton's group at Boston College developed into the TIMSS International Study Center. The center became responsible for the subsequent TIMSS studies and the PIRLS (i.e., reading literacy) studies and is now known as the IEA TIMSS and PIRLS International Study Center.

TIMSS— the IEA's third study in both mathematics and sciences-- was such a success that IEA decided to make it part of its regular cycle of studies (see above and Figure 1). To keep TIMSS as the 'brand name' for the studies in both subjects, the IEA needed to find another meaning for the acronym, and so from the mid-1990s onwards , TIMSS has been known as the *Trend In Mathematics and Science Studies*. Since the third assessment of mathematics and science in 1995, the IEA has conducted TIMSS-1999, TIMSS-2003, TIMSS 2007 and 2011.

### **Creation of the IEA Data Processing Center**

At the 1988 GA meeting (Frascati, Italy) it was decided that the new IEA Headquarters should provide, amongst other things, "handling of all international data processing and preparation (e.g., manuals, code-books, etc.)". This is an interesting statement, if one reflects on a few developments that took place in the IEA in preceding years.

#### ***The 'first' IEA International Data Center***

I think that not many IEA members at the 1988 GA realized that in 1986 a contract had been signed between the IEA and SVO, the Netherlands Foundation for Educational Research arranging:

the creation of an International Data Center (IDC) handling qualitative and quantitative research data stemming from comparative projects carried out in many countries by the IEA. This appears to be a promising way to provide optimal use of

expertise, knowledge and other resources, all of which should yield rich dividends in national and international contexts since it is our view that national educational policies cannot be usefully developed without explicit knowledge about education in other countries. ( Preamble of the contract).

The contract further stated, as aims and objectives of the IDC, that the center “will be responsible for the cleaning, editing, sorting and other statistical analyses (including the calculation of standard errors of sampling) for all national data sets IEA deems relevant”. The IDC would create a data archive, promoting and facilitating its use, and would also provide consulting services and knowledge and skills-transfer activities to individuals chosen by SVO or IEA. It was also agreed that the IDC would start with a ‘trial period’ of one year (September 1, 1986, until November 1, 1987) during which the feasibility of establishing the center would be explored.

The contract was signed on August 29, 1986, by the IEA Chair, Alan Purves, and by representatives of the SVO. Considering the date of signing with the fact that Neville Postlethwaite held the IEA chair until the 1986 GA meeting, I assume that most of the preparatory work for this International Data Center took place under Postlethwaite’s authority.

The minutes of the 1988 GA meeting in Frascati make no reference to this IDC, nor do I recall this center being mentioned during the discussions on IEA’s future at this meeting. So i believe we may safely conclude that the first IEA International Data Center never took off!

### *Developments since 1988*

In the discussions at the 1988 GA meeting about IEA’s future, one important issue concerned the extent to which the IEA would need a centralized versus decentralized structure. In his report on this GA meeting, Jan de Jonge, the SVO observer, wrote that Postlethwaite advocated greater centralization, and therefore wanted one International Data Center. De Jonge also said that Postlethwaite had

lost 'the battle' despite the contract between the IEA and SVO ( see above). The most important argument of those favoring decentralization was that expertise in the subject domain of a study and expertise in the methodology should not be separated, and as it seems not feasible to have one IEA coordinating center with domain expertise in all studies, they believed that IEA should not centralize the coordination of its studies like when it conducted the 'Six-Subjects Study'. The minutes of this 1988 GA meeting (p.5) state that

- (a) The original centralized structure for all studies was changed in 1974 as it became too difficult for one center to administer all studies.
- (b) This issue should now be re-examined as a means of more controlled conduct of studies and for a more organized/unified approach to fundraising.

As a consequence, the IEA decided in 1988 that the new Headquarters should be in charge of handling all international data processing of the studies.

As presented above, the proposal for the IEA Secretariat listed a number of tasks in the area of data management that could be seen as a statement of mission and purpose for the Secretariat's activities in the area. But it took a few years before the IEA was in a position to formulate concrete plans to shape this task.

In January 1993, Andreas Schleicher, who was until then responsible for data management at the ICC for the Reading Literacy Study, joined the IEA Secretariat as Director of Data Management and Analysis, tasked to organize the data management function of the Secretariat. His appointment took place in the same period that TIMSS was reorganized and the TIMSS International Study Center was established at Boston College. But his appointment also coincided (almost) with the closure of the Reading Literacy (RL) Study. Given the excellent work conducted by the data management group at the ICC for Reading Literacy in Hamburg, it was decided to keep this group operational,



and in 1993 the GA (El Escorial, Spain) passed a motion that the Secretariat would develop a technical unit named the IEA Data Processing Center (DPC), functioning at the German IEA office at the University of Hamburg. This was possible as the German IEA office was linked to the group of Rainer Lehmann, the German GA representative and colleague of Neville Postlethwaite at this university. Formally, the DPC acted under the auspices of a private foundation called 'IEA Hamburg e. V<sup>4</sup>.'

The activities of the Data Processing Center in its first year as an IEA center (August 1993 - August 1994) involved: providing technical and operational support for TIMSS; establishing the IEA Data Enhancement Project (initiated by IEA to standardize data archiving - including past projects - to ensure easier access and dissemination); creating the Reading Literacy archive. At the same time, the Center also worked for other international organizations - such as OECD, the World Bank and the IIEP (UNESCO). The 1994 GA meeting congratulated the DPC on the Center's progress and operation and *expressed the desirability of maintaining a central data processing center.*

One may wonder why this notation of the "desirability of maintaining the DPC" was needed, as it would seem unnecessary considering the high praise given to the DPC. The explanation is that Neville Postlethwaite was about to retire from Hamburg University whilst, almost simultaneously, Rainer Lehmann had accepted a position at Humboldt University in Berlin. This university was to be the future German IEA office, and although it would assume responsibility for national membership dues, it did not seem to have the possibility of hosting the IEA Data Processing Center as well. Thus, the relationship between IEA and the University of Hamburg was suddenly severed.

Without the 'shield' of the German IEA office,, the IEA DPC could not continue as a unit within the University of Hamburg. The dilemma

---

4. e. V. means "eingetragener Verein" , i.e., registered non-profit organisation

was clear: the IEA had finally a well-functioning DPC, but – as it was going to lose its formal position as part of the University of Hamburg – there was a real threat that the DPC could no longer keep its non-profit status and, as a consequence, would have to be closed down.

I remember this dilemma vividly– I also recall that in my capacity as IEA Chair I did not want to lose this center because in its short period of functioning it had already proven itself so invaluable. Although I have no records of a meeting with the SC at that time, as I seem to recollect, I convinced the SC that the DPC should be maintained and that this could be achieved by making it formally part of the IEA Secretariat under direct administrative assignment to the IEA Executive Director. The SC agreed with me because - although there was no 'business plan' - the DPC's immediate future was secured because it was needed for various tasks for TIMSS, such as providing technical and operational support for the field operations data entry and data management, which would pay the costs to keep the center up and running for the first few years. The 1994 GA approved the SC proposal to make the DPC formally part of the IEA Secretariat, i.e., its staff becoming IEA personnel.

As of 1995, the DPC staff comprised four professional staff members, supported by graduate students employed on a temporary basis, co-directed by Heiko Sibberns and Dirk Hastedt. At the end of my tenure as IEA Chair (in 1999), the DPC had 13 full staff members and 15 assistants (supported by about 50 coders).

In retrospect, the decision in 1994 to formalize DPC's function and status in the IEA appeared to be a very good one. Since that time the IEA DPC has been responsible for all data processing related activities of all IEA projects. This has included all TIMSS cycles (1995, 1999, 2003, and 2007), all PIRLS cycles (2001, 2006), the CIVIC and ICCS study, the TEDS study, and the SITES study. In addition, in 1998, the DPC also took on the responsibility for conducting large-scale assessments in Germany, including state comparisons for national educational stan-

dards, and, more recently, the National Educational Panel Study (NEPS). This responsibility began with the planning and data collection of OECD's PISA study in Germany. Ever since that time, the DPC has been responsible for the data collection of all PISA cycles in Germany.

Today the IEA DPC is responsible for conducting nearly all IEA studies in Germany, through subcontracts from universities or the ministries of education. This enables the DPC staff to gain experience similar to what research centres in other countries experience when conducting IEA studies in their countries and helps them to understand and solve challenges that participating countries might face. Since 2007, the DPC has also been functioning as a data processing center for non-IEA studies, such as the OECD's Teaching and Learning International Survey (TALIS) or OECD's Programme for the International Assessment of Adult Competencies (PIAAC). For the OECD TALIS, the DPC has also been appointed as the international study center.

In addition, in 2007 the DPC increased its research activities by establishing a new unit, the research and analysis unit (in collaboration with the Education Testing Services, ETS, in the USA). Since that time, the acronym DPC has changed to stand for "Data Processing and Research Center".

### **Towards a Publication Policy**

One of the challenges confronting the IEA in its transition from a researchers' organization with studies driven purely by researcher interests to a professional organization with a structure and program of studies suitable for meeting the growing interest of policy makers and practitioners, was the need to improve the publications of the research findings.

The dilemma was clear. As an organization of researchers, the IEA was

most interested in publishing their research data, and at this time (the late 1980s), the IEA was not fully aware of policy makers' need for up-to-date information on the achievement of their schools and students. The problem can be illustrated with the example of the SIMS (Second International Mathematics Study) publications: the data collection for this study took place in the northern hemisphere in the spring of 1981, but it was announced at the 1988 GA meeting that the first two of the three volumes of this study would be published in 1989. This slowness was of course partly a consequence of the lack of funding, as SIMS officially ended in 1986, and further work on the research volumes had to be done by the authors in their own time. But a more important reason was that studies like SIMS were designed as research studies with the intention to publish only research volumes. At that time (late 1980s) the IEA had an agreement with Pergamon Press for the publication of the research volumes, and besides these volumes, only national reports were published (e.g., the Dutch national SIMS report in 1983!). I use SIMS here as an example, but the problem also occurred in other studies as well.

The discussions at the 1988 GA meeting showed that the membership was aware that a reorganization of the IEA should also include the timely publication of findings. This was evident in statements such as: (i) policy-relevant publications should appear before the end of a study; (ii) the number and type of reports to result from a study should be included in the planning and the design of the study; (iii) an international report should be published within one year after the international centre had received the data; and; (iv) reports should stress information/ analyses that cannot be obtained from a national study (taken from the 1988 GA minutes). Interestingly, one other issue was expressed by the membership: the concern that early 'short' reports that did not include interpretation would leave out too much of the real situation and would end up focusing on country rankings.

But the message about timely publications was clear and well understood. For example, in the proposal for the Third International

Mathematics Study (TIMSS) discussed at the 1989 GA meeting in Seoul, David Robitaille stated that several deadlines would be firmly established from the outset: (i) clean, national datasets were to be submitted by the National Centres to the ICC within 6 months of the deadline for completion of data collection; (ii) publication of the international report of descriptive analyses and preliminary results of the study would take place within 12 months of the deadline for completion of data collection; and (iii) the databank would be available for secondary analyses within 18 months of the deadline for completion of data collection.

The SC was convinced that the IEA should change in this respect. I could state in my first chairman's report to the IEA (GA meeting of 1990, Beijing) that:

The regaining of the confidence in IEA is based on the expectation that IEA will be a more centralized monitored and administered organization. In the many conversations about the 'new IEA' during the courtesy visits I brought this year to international and national agencies, it was clearly and explicitly expressed that this centralization of IEA was considered to be a necessary condition for cooperation with and giving credit to IEA.

Let me illustrate this with two examples. First, most donor agencies are not interested in our research volumes, but in quick results, published about a year after data collection in an understandable format. So far, IEA studies have not proven being able to meet this kind of expectation. It is my experience that the Standing Committee and Headquarters (if not the chairman in person) will be held accountable by donor agencies for keeping promises about delivering results in time and in a proper format.

Secondly, IEA studies which will only address interests of researchers and educators, and which are not open to the questions of policy makers, can no longer expect substantial financial support from governmental agencies. Therefore, IEA

studies must balance the needs of researchers, practitioners, and policy makers.

An illustration of this attitude was that in my communication about 'TIMS becoming TIMSS', the NCES indicated that it expected high-quality survey practices combined with timely release of data reported in a readily understandable format 12 months after data collection. The 1990 GA meeting (Beijing) discussed these issues, which resulted in the IEA Publication Policy and Procedures, published in the 1991 IEA Guidebook. Here a distinction was made between two types of international publications documenting international comparisons; *Study Volumes and Short Reports*, elaborated as follows (1991 IEA Guidebook, p.129):

*Study Volumes*: in depth descriptive and/or explanatory analysis of major issues/findings (200-300 pages in length) published at the conclusion of each IEA study. These reports tell the results and educational implications of each study and are aimed at educational policy makers, administrators, university and teacher training college personnel. (...)

*Short Reports*: brief 80-100 page booklets highlighting the major comparative outcomes of a study alongside basic contextual information. The report is used to inform the educational community worldwide. Additional booklets may be written on topics addressing unique issues associated with a particular study. The first short reports will be published within 12 months of completion of data collection. Short reports may not necessarily contain results from all participating countries. If countries are very late in sending in acceptable data files, they may not be included in the preliminary report.

All the IEA studies with data collection in the 1990s tried very hard to meet this policy and were, in fact, able to do so. For example, the *Reading Literacy Study*, with data collection in the period October 1990 to April 1991, published its first short report of 87 pages (+ Appendices) in July 1992, followed by a second one in November 1992 and a third one in January 1993, whilst a report with explanatory

analyses appeared a few years later (in 1996). The Computers in Education (*COMPED*) study, with data collection in 1989, published its 'quick report' in early 1990 and a research volume in 1993. Finally, TIMSS had its data collection in late 1994 (southern hemisphere) and May and June 1995 (northern hemisphere) and the reports were published in November 1996. In sum, one may conclude that this aspect of IEA restructuring has also been successfully implemented.

### **Some Final Reflections**

When I took on the chairmanship of the IEA, I was given the task of providing leadership in the process of transforming the IEA into a well-structured, professional organization. The IEA General Assembly in 1988 (Frascati, Italy) and the Standing Committee (SC) in its meeting in February 1989 decided on the direction of the change and identified the priorities. These included establishing Headquarters (or a Secretariat) and developing a membership and fee system to ensure that the international costs could be covered, as well as addressing a number of aspects crucial for realizing a successful transformation, including the development of a program of studies (with special attention on how TIMS changed into TIMSS), the creation of the IEA Data Processing Center (DPC) and the development of a publication policy. In retrospect, I conclude that the IEA Standing Committee (in its various compositions – see below), in those first years of this transition process, provided creative, constructive and strong leadership. My preceding description of how we addressed the various components that needed to change illustrates that the IEA members – although they were often quite critical in the discussions about the changes proposed by the SC – actually functioned as a group of 'critical friends' who contributed in a constructive way to the process of transformation.

It is not surprising then, that the revised IEA appears to be a dynamic organization. I hope that in the near future the IEA will demonstrate that it is willing to serve the needs of educational practices and poli-

cies and the monitoring thereof in a dynamic and constantly changing world. I next will present a few examples (without trying to be complete and without knowing the status of IEA's thinking on them) to illustrate areas for further growth and development that I think are not only promising, but also very important:

- (i) The IEA has expressed its support for the project 'Assessing and Teaching 21st Century Skills' (see: <http://atc21s.org/default.aspx>). I hope that the assessment of such skills – also called 'lifelong learning skills' – will become an integral part of IEA's assessments.
- (ii) Increasingly, countries (or education systems) are developing national systems to monitor educational quality. The relevance of IEA studies will increase when educational systems have the opportunity to include national options with IEA tests, in order to link their national assessments with international ones.
- (iii) Related to the previous point, IEA's experience and expertise can be utilized to offer its members the possibility of conducting in-depth diagnostic secondary analyses of their IEA data aimed at identifying causes of deficiencies that appear from the descriptive comparisons.

Today I am –and for many years now have been– an 'outsider' to the IEA and its studies. But I consider it a privilege to have had the opportunity to chair the IEA for almost ten years. A personal and very enriching benefit of this period has been the friends that I made from all around the world, and who will be my friends for a lifetime!

Finally, I want to acknowledge the contribution of all the people who collaborated with me during my tenure as IEA Chair:

- members of the Standing Committee, in the order that I worked with them: Alan Purves (USA), Zoltan Bathory (Hungary), Gilbert De Landsheere (Belgium), Robert Garden (New Zealand), Inger Marklund (Sweden), Moegiadi (Indonesia), Shin Se-ho (Rep of Korea), Leslie McLean (Canada), Gordon Ambach (USA), Alejandro Tiana



- Ferrer (Spain), Marit Granheim (Norway), Ryo Watanabe (Japan), Georgia Polydorides (Greece), Hans Wagemaker (New-Zealand), Christiane Brusselmans-Dehairs (Belgium), David Nevo (Israel), Peter Vari (Hungary), Armin Gretler (Switzerland), Ester Ogena (Philippines), Constantinos Papanastasiou (Cyprus) and possibly others;
- staff of the IEA Secretariat, especially Bill Loxley (1990 – 1994), Barbara Malak (since 1995), Hans Wagemaker (since 1997), and also Wim Hayes, Leendert Dijkhuizen and Jur Hartenberg (who provided me with the data on finances);
  - international coordinators of the studies, especially Neville Postlethwaite, Hans Pelgrum, David Robitaille, Al Beaton, Judith Torney, Ina Mullis and Michael Martin.

## References

For this chapter I used as reference material the minutes of the various meetings of the General Assembly, as well as a number of documents that were part of the documentation for those GA meetings.

Among the IEA documents that I referred to are:

Burstall, C. & Marklund, I. (1988). *Discussions in General Assembly concerning the future of IEA*. Memorandum prepared for the GA meeting in Frascati (Italy).

Netherlands Foundation for Educational Research, SVO (1989, March 20). *Proposal for housing the IEA-Secretariat at SVO in The Hague*, The Netherlands

Plomp, T. (1990, September). *IEA in transition*. Report of the IEA chairperson to the 31st IEA General Assembly, Beijing.

Further:

President George H.W. Bush, (January 31, 1990). *State of the Union*.  
<http://www.infoplease.com/t/hist/state-of-the-union/203.html>  
 (retrieved from the Web on May 22, 2011)

## CHAPTER 3

# **The Second International Mathematics Study (SIMS): Intention, Implementation, Attainment**

*Kenneth J. Travers*

### **Prologue**

From my vantage point as Chairman of the International Steering Committee, I share some of my (filtered!) recollections of SIMS: key personnel, major events, noteworthy outcomes. The chapter has three major parts. Sections 1-4 address SIMS Intentions – the goals, conceptualization and plans for the project. Section 5, Implementation, deals with SIMS activities, focusing on the centrality of the national center in each country in helping to make sure that the project took place in ways that were intended. The chapter closes (sections 6-7) with SIMS attainments – reflections on what I see as some of its major outcomes. As reiterated throughout, it is the people who make IEA projects happen: those who have the vision as to what needs are to be addressed, and how to go about doing so; those who have access to means for funding (extremely important!) and those who actually do the work. SIMS took place in critically important times. In the 1960s and 1970s there had been mathematics curriculum reform in the schools of the participating countries. Arguably, the SIMS findings helped to form the basis for needed steps in revitalizing the teaching and learning of mathematics in the following decade and beyond.

### **Beginnings of SIMS**

My introduction to cross-national empirical studies of education was unexpected. That this introduction happened to be in mathematics education was an amazing bonus. In 1966-1967 I was spending a post-doctoral year in curriculum studies at Stanford University under the

direction of Lee J. Cronbach and Edward G. Begle and learned about a campus-wide lecture by one Torsten Husén. As it turned out, the late and much revered Professor Husén was at the Center for Advanced Study in the Behavioral Sciences, next door to Stanford. His lecture impressed me greatly, first because it dealt with a historic empirical research project, which for those days was large scale (TWELVE countries!) and taxed the limits of computer data storage (visions of a room full of boxes of punched cards!). Furthermore, the IEA was pursuing research implications of the daunting notion of 'the world as an educational laboratory'. As Husén stated:

In general terms, international studies such as this one can enable educationalists (and ultimately those responsible for education planning and policy making) to benefit from the educational experiences of other countries. It helps educationalists to view their own system of education more objectively because for the first time many of the variables related to educational achievement had to be quantified in a standardized way. (Husén, 1967, p. 13-14)

It was about eight years later when I was back at Illinois that my then colleague, the late and honored Alan Purves informed me that an 'IEA person from Wellington' was visiting the campus to explore the possibility of there being a SECOND International study of mathematics, and asked if I would like to meet that individual. I was of course interested in meeting Roy W. Phillipps, who was the Executive Director of IEA at the time.

An initial step in establishing SIMS was the appointment of Roy Phillipps as Chairman of the International Project Council. The International Project Coordinator for SIMS was to be Robert A. Garden. Bob, an experienced and highly able specialist in mathematics testing, was in the New Zealand Ministry of Education.

It was a great personal privilege for me to get to know Roy. He was unassuming, perceptive (and as I soon learned), persistent! Some time later that semester, Alan Purves let me know that IEA had invited me

to chair the International Steering Committee (which soon came to be known as the International Mathematics Committee) for SIMS and Roy would be returning shortly to the United States to join me in the search for funding.

I quickly came to appreciate Roy. For the next several months I was his apprentice as he and I made our way around Washington DC and New York City in search of financial support for SIMS. During the spring of 1976, things moved quickly. Roy spent countless hours in the air between New Zealand and the United States.

Sadly, Roy has passed away. I am confident that it is the hope of all of us who served on the International Mathematics Committee that SIMS in general, and this chapter in particular, are fitting memorials to him and his enormous effort in helping to make the project a reality. In fact I would love to regard this chapter as a 'Mini-Festschrift' for Roy W. Phillipps.

In early 1976, the US National Institute of Education commissioned James Wilson at the University of Georgia to write a paper on the topic (roughly), 'If a Second International Mathematics Study were to be carried out, what its goals should be'. In part, Wilson (1976) stated,

...the second study should be approached as a curriculum survey. If the study focuses on the content of what is being taught, the relative importance given to various aspects of mathematics, and the student achievement relative to these priorities and content, then the international and national results can help in our understanding of comparable curriculum issues.  
(Mimeo)

As it turned out, the Wilson document played a critically important role in helping to formulate the goals for SIMS.

### **International Meetings 'Unlimited'**

The first of many meetings of the International Mathematics Committee to plan, prepare instruments, do pilot studies, monitor

data collection and analysis, and of course prepare international reports, was held in St. Andrews, Scotland, in June 1976. The members of the Committee (See Figure 1) were: Sven Hilding (Sweden), Gerard (Gerry) Pollock (Scotland), Hans-Georg Steiner (Germany), Fred Van der Blij (The Netherlands), Tamás Varga (Hungary), together with James Wilson and Edward (Skip) Kifer, both of the United States.



**Figure 1.** The SIMS International Specialist Committee:

**Back row:** Fred Van der Blij; Sven Hilding; Edward (Skip) Kifer; Bob Garden  
**Front row:** Ken Travers, A. I. (Izzie) Weinzweig; Roy Phillipps.

**Not shown:** Gerry Pollock, Hans-Georg Steiner, Tamás Varga, James Wilson, Richard Wolfe.

This first meeting was memorable for several reasons. Our hosts, the Scottish Council of Education, under its director, Bryan Dockrell, accorded us every courtesy. The setting (even for those of us not golfers...smile) was spectacular. And here the basic elements of SIMS were laid out. A presentation by David Walker, the well-respected IEA researcher, helped to link us with key assumptions and priorities in educational research.

It was decided that SIMS would target two populations for study, par-

alleling those for the First International Mathematics Study:

*Population A: All students in the grade in which the modal number of students has attained the age of 13.0 – 13.11 years by the middle of the school year.*

*Population B: All students who are in the normally accepted terminal year of the secondary education system and who are studying mathematics as a substantial part (approximately 5 hours per week) of their academic program.*

The Committee very much wanted to include a younger age group (for example, nine year olds) as well. However, the time and expense involved in breaking new ground with such a population soon led us to drop this idea. In retrospect, this was indeed a wise decision.

Fortuitously, the Third International Congress on Mathematics Education (ICME-3) was scheduled for West Germany in August of the same year. We decided that it was strategically important to present preliminary ideas for the Second Study in an informal session at the Congress in order to help alert the international community that SIMS was underway. Concurrently, the Mathematics Committee recommended to the IEA National Centers that they provide information about SIMS to their respective national communities of researchers and educators.

Let me say right away that the International Mathematics Committee, together with our consultants, were absolutely terrific colleagues. Not surprisingly in an undertaking of the scope of SIMS there were times of 'tough sledding'. But through it all, the Committee worked smoothly. Perhaps the intervening years have helped to embellish my memory, but I do not recall even one occasion during our many sessions in diverse cultural contexts that there was ever a harsh word among us. Truly remarkable.

*Appointment of SIMS consultants:* In view of concerns about cross-national empirical research in mathematics education that had been expressed by various scholars (for example, Freudenthal, 1975), the International Mathematics Committee took such considered commentary into account in the design of SIMS. To help guide us we recruited

additional specialists including psychometricians Richard G. Wolfe, Ontario Institute for Studies in Education, Canada and Leigh Burstein of the University of California, Los Angeles, USA. Furthermore, A. I. (Izzie) Weinzweig, Professor of Mathematics at the University of Illinois at Chicago was recommended by the US National Institute of Education to serve as a consulting mathematician.

*Parenthetical note:* Even though the International Mathematics Committee, as well as staff resources of national research centers, provided ample expertise in mathematics test development, the wisdom of having an expert mathematician on hand was affirmed on several occasions, one especially memorable. When the SIMS findings were released, they were featured in the US on the front page of the *Sunday New York Times*. Because of the emphasis placed by the Committee on the content of the curriculum (in keeping with the admonitions of the Wilson paper!), we provided illustrative items with the summary statistical data.

On the Monday after the release, I got an urgent message from the *New York Times* Education Editor stating that someone had called in to say that there was a wrong answer for one of the items. I feverishly checked the item and it looked fine. But for insurance, I immediately contacted Izzie Weinzweig. Fortunately, all was well. Verily, those of us in the testing business must learn to live with such 'nightmare scenarios'.

On the lighter side, concerning the travails of test development, very early in SIMS we were looking for schools in which to try out the survey items. On one occasion, I called up a colleague at a remote campus and asked if she could help us by finding some classrooms. She replied, "*Ken, we are so isolated here I would gladly pay for any phone charges to reach me!*" Needless to say, this site was one of the most responsive and helpful ones in which we worked.

On a related note, at another location, we were field testing the student background/attitude surveys. One item read, '*Mathematics is a good field for creative people*'. In going over the responses we found the fol-

lowing, in carefully scripted youthful lettering on the survey page itself: *'How would I know?'*

## Gathering Support for SIMS

Our Dean of Education at the University of Illinois, J. Myron Atkin, was extremely enthusiastic about a Second International Mathematics Study and committed funding to convene on campus key mathematics educators from around the US in order to gauge national interest in the project. The conference was held at the University in 1977, with the Wilson paper again helping to frame its goals. Roy Phillipps was the featured presenter, setting the tone of the meeting and outlining priorities for such a study. The sequencing of events seemed to Roy, and to me, important in terms of seeking financial support. Since the sources of funding for the international aspects of SIMS were most likely to be in the United States, it was essential in talking to potential funders to be able to unambiguously state that there was strong support for the project among the key US constituencies, especially mathematicians and mathematics educators.

The conference was memorable. Many of us were filled with a sense of awe as it became clear that we were on the verge of engaging in an extremely ambitious empirical cross-national study of the teaching and learning of school mathematics. A noteworthy moment was in the opening session when Roy Phillipps, in his self-effacing manner, quipped that as he pondered the enormity of the task facing us, he was reminded of the first lines of the hymn:

*Turn back, O man, foreswear thy foolish ways.*

As I was soon to learn, such a remark was vintage Roy.

In terms of funding, substantial support for international costs was eventually provided by the US National Center for Education Statistics, with special interest expressed by Director Emerson Elliott and Program Officer Larry Suter. The US National Science Foundation,



under leadership from Program Officer Ray Hannapel, also provided support at critically important junctures.

In retrospect, it is clear that 'the stars were being aligned' in favor of a second mathematics study. Important impetus was added in the early 1980s, when the *Nation at Risk* (1983) report by the US Department of Education placed emphasis on shortcomings of the country's educational system from an international point of view. The report begins with this ominous declaration:

Our Nation is at risk. Our once unchallenged preeminence in commerce, industry, science, and technological innovation is being overtaken by competitors throughout the world. (p. 1 )

As a part of the resource documents for *A Nation at Risk* I was commissioned (Travers, 1983) to provide a summary of the goals and current accomplishments of SIMS. The *Nation at Risk* report served in the US to heighten awareness of the nation's needs for educational reform. The prospect of the availability of SIMS data with its focus on the centrality of curriculum and instruction helped greatly in attracting the interest of funding agencies in much needed ongoing support for both the international and US national components of SIMS.

The second Meeting of the International Committee was held in 1977 at Shakertown, Pleasant Hill, Kentucky USA. The conference site is a reconstructed historic village in the picturesque Blue Grass countryside. The meeting was supported in part by the University of Kentucky College of Education. The purpose of this meeting was to prepare a concept paper that outlined the major goals, design, and proposed instrumentation for SIMS. The paper was needed for consideration by the IEA General Assembly when it met in Tokyo. The respected, late T. Neville Postlethwaite, then Chairman of IEA, was in attendance, providing guidance as to expectations of the General Assembly for SIMS.

*IEA General Assembly, Tokyo, January 1978:* The Assembly provided many of us on the International Mathematics Committee the first

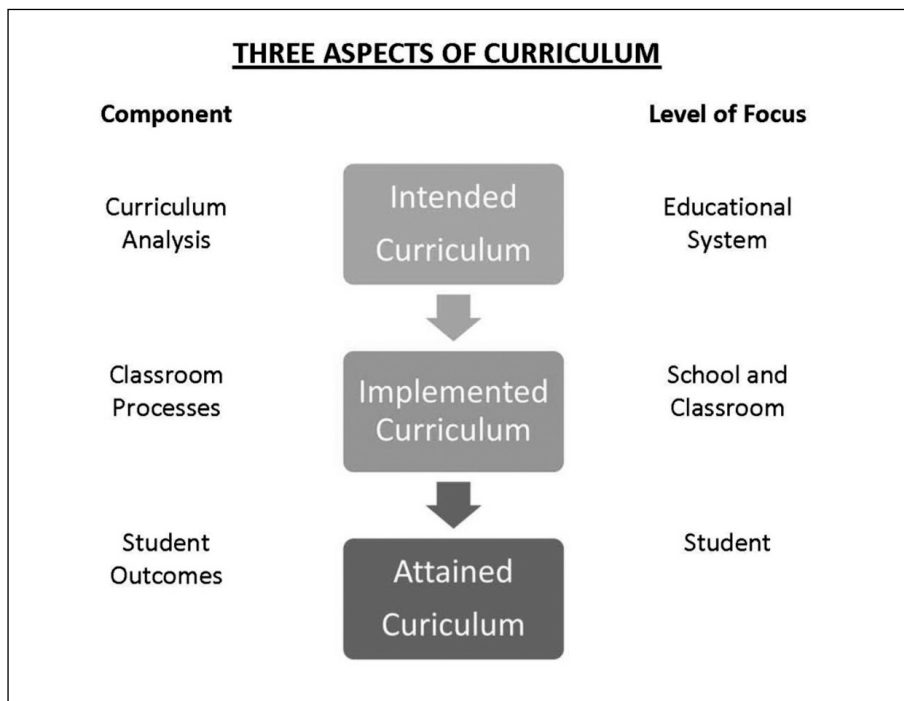
opportunity to meet the major figures in IEA. Also attending were the National Research Coordinators who had a direct interest in SIMS. At that meeting, I learned firsthand what it is that makes IEA 'tick'. Their extremely ambitious comprehensive projects require enormous commitments of national research centres and individuals around the world. Our gracious hosts were IEA Council Member, Hiroshi Kida, and National Research Coordinator, Toshio Sawada.

On the first day of the Assembly, it happened that on page one of Tokyo's English language newspaper there was a photograph of young people in a reverent pose. The caption read, 'High school students pray for success on national examinations'. This was powerful testimony that cross-national variation in cultural values does indeed play a significant role in accounting for differences in academic achievement!

### Conceptualizing SIMS

The overarching concern of SIMS for the curriculum led to our conceptualization of what came to be known as 'The tri-partite curriculum' (Figure 2):

- (i) The content **intended** to be learned: We obtained this from a topic by topic cross-national synthesis of the mathematics curricula for the two target populations (roughly, 13-year-olds and the end of secondary school cohort, respectively). The data were based on fine-grained content grids that we developed and sent to each national center in order to provide national snapshots of what mathematics was intended to be learned in each participating country.
- (ii) The content that was **implemented**: Our goal here was to get data at the classroom level on what mathematics was taught by the teacher. This measure of the implemented curriculum was derived from the Opportunity to Learn (OTL) data referenced below.
- (iii) The content that was **attained** (learned). This information came from the SIMS tests taken by the students.



**Figure 2.** The ‘tri-partite’ view of the curriculum provided a basis for the design of SIMS, guiding instrument development and data collection. This formulation was used by IEA throughout all subsequent studies.

The tri-partite perspective of the curriculum provided a road map for the design of SIMS and the preparation of the three international reports. As Plomp (2011) has noted (personal communication) this formulation was used by IEA throughout all subsequent studies.

### **The Centrality of the IEA National Research Coordinators**

As I have already remarked, the National Research Coordinators (NRCs) play essential roles at each level of the project: Intention, Implementation and Attainment. Consequently, regular meetings of the NRCs were critically important. Fortunately, each of the several successive meetings of the NRCs was extremely productive—a fact

made possible primarily by the personal commitment of the respective coordinator and his/her personal staff, as well of the host institution. Furthermore, of course, there was sustained work by the International Committee members and the project staffs 'back home' in Wellington and in Illinois. I will not here give a full account of each meeting. However, I reiterate that each and every get-together was an indispensable link in the chain of events leading to the productive and informative outcomes of SIMS. As I reflect on those many sessions, I am once again impressed by the reliable work of each researcher and of the respective national centres to the improvement of teaching and learning in their nation's schools.

### *Dobogókő, Hungary*

One of the more memorable meetings was in the late 1970s in Dobogókő, a scenic conference site in the Hungarian countryside. Our hosts were IEA Council member Zoltán Báthory and National Research Coordinator Julia Szendrei. My introduction to Hungary was rather stark (recall that this was prior to the falling of the Berlin Wall). Sven Hilding and I had chosen to travel by train from Paris to Budapest. The customs/immigration procedures at the border were extremely rigid. Security officers did painstaking searches, both within our compartment and outside, underneath the railroad carriage. Interestingly, such strict inspection protocols were not as evident at the airport upon our departure.

We were transported to the meeting site along an isolated road. During the trip we stopped suddenly. It was startling on looking outside to see a checkpoint manned by Soviet soldiers. One of our party seemed especially on edge. He later let us know that he was in the US army reserve!

The Dobogókő sessions were excellent, thanks mainly to our hosts who ensured that everything went smoothly. And the cuisine was of course superb.

I also recall a subsequent meeting that took place in Budapest. At a reception, International Committee member Tamás Varga remarked to me, *'Ken, a colleague of mine here in Hungary has developed a unique, hands-on puzzle. He would like to know if perhaps there would be interest in this 3-D brain teaser in the United States. My colleague's name is Rubik'*. Tamás then handed the device to me for a brief inspection. Some time (perhaps one year) after my return to the US, Rubik's Cube appeared on the market. Hence (admittedly without validating data!) we might claim that SIMS helped facilitate the introduction of the Cube to North America!

### *'The curriculum as a variable': Theme of German Symposium*

The emphasis in SIMS on the curriculum led to a symposium hosted by Hans-Georg Steiner, Institute for Mathematical Didactics, (then) West Germany. The sessions were at a Franciscan conference center, Haus Ohrbeck, situated in the foothills of the Teutoburger Wald. Extensive preparations for the meeting were made by Hans-Georg Steiner and by my Illinois colleague, Ian Westbury, who was then editor of the internationally respected *Journal of Curriculum Studies*. Westbury (1980) prepared for the conference a comprehensive set of papers *'Change and Stability in the curriculum: An overview of the questions'*.

The sessions were extremely productive in mapping out curricular components to be addressed by SIMS. One vivid recollection I have of the Haus Ohrbeck Center is that promptly every morning the sisters would briskly whisk through our dormitory rooms, ensuring that all was in order, and energetically throw open the windows in order to take advantage of the crisp January air. In retrospect, it's not surprising that with such a vigorous introduction to the morning all conferees were promptly seated in the meeting hall for each day's beginning session!

At this point, I would like to acknowledge the extraordinary work done by Ian Westbury on other aspects of SIMS, as well. It is fortunate for SIMS that he committed a great deal of personal effort to archiving the SIMS data. Thanks to Ian, together with our other colleagues at the

Illinois and New Zealand Centers, the data were made secure and accessible for secondary analyses.

## SIMS Reports

Three main reports were produced, as briefly referenced below. The first volume gives background for SIMS, and then provides findings for the **intended** curriculum (cross-national patterns of mathematics content of school curricula). Volume I also reports on the **implemented** curriculum (teacher ratings of student opportunity to learn the curricular content that was on the test). Volume II addresses other major outcomes, with an emphasis on the **attained** curriculum—student achievement data. In Volume III there are detailed accounts of findings of the ‘longitudinal study’, an optional version of SIMS that focused on student growth in achievement during the school year and on associated instructional practices used while key mathematics topics were taught and learned.

### *Volume I: Analysis of Mathematics Curricula* (Travers and Westbury, eds., 1989)

A major component of Volume I is the detailed documentation of cross-national variation in curricular content. One important finding is that as a result of curriculum reform initiatives in the 1960s and 1970s, there was substantial variation, especially in geometry, in the intended curriculum. To be sure, these differences led to challenges in test construction, for as Husén (1967) remarked concerning the First Study:

With some justification, one might paradoxically say that the tests devised for the IEA study are equally appropriate or inappropriate to all the systems participating in the study. (Vol 1, p. 21)

### **Euclid must go!: The quest for the equally fair and unfair test**

As we report in Volume I the influence of the Bourbaki group of mathematicians, especially those in France and Belgium, led to a distinctive

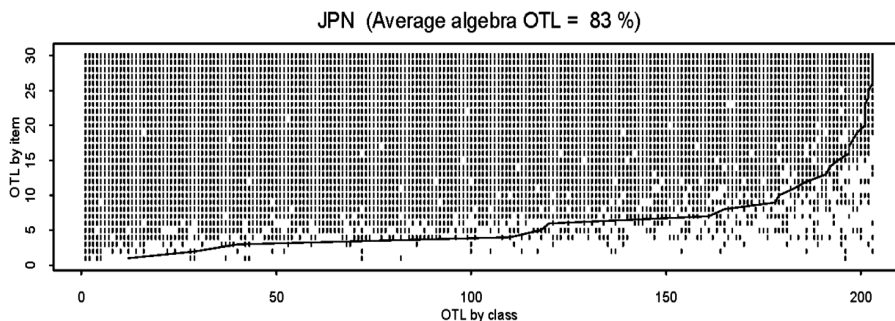
view of school mathematics, characterized by the slogan 'Euclid must go!'. As noted by Servais (1975) the Bourbaki program outlined an approach to geometry that was based on a vector space of two dimensions. The result was a dramatically new curriculum which our SIMS data suggested made inroads in the curricula of only a few countries. These differences raised substantial challenges when the tests were developed. However, with persistence, and the patiently pervasive (and at times seemingly limitless!) good will of all National Research Coordinators, the SIMS tests were finally in place, characterized perhaps by the result of any successful negotiation: all parties were reasonably satisfied, with each one having lingering residual concerns about the final product.

Volume I also deals with the Implemented Curriculum. Here, we pay considerable attention to the construct 'Opportunity to Learn' (OTL). As already noted, there had been concerns expressed about the methodology of the First Mathematics Study, especially about the opportunity to learn measure (Freudenthal, 1975). After much deliberation and some trial field explorations, we agreed on revised wording for OTL for the teacher questionnaire that said, in part, for each item on the SIMS tests (See Travers & Westbury, 1989, p. 112):

During this school year, did you teach or review the mathematics needed to answer this item correctly?

As we considered ways of effectively analyzing and reporting OTL, we were assisted greatly by the expertise of Richard Wolfe at OISE, Toronto, Canada, who devised a way of analyzing the data that draws heavily on the analysis of patterns of achievement that uses Student-Problem (or S-P) charts (Harnisch & Linn, 1981).

In SIMS, the S-P approach was used for teacher OTL data instead of student achievement data. The result is called an Item-Classroom (I-C) chart. To illustrate, see Figure 3 which presents OTL data on Algebra for Population A students in Japan (Note: In Japan the students were 12-year-olds instead of the internationally defined 13-year-old population).



**Figure 3.** I-C chart for Algebra, Japan, showing rather full teacher coverage of the Algebra items on the SIMS test.

*Note:* The data suggest here (and are confirmed by components of variance analysis) that Japan did very little, if any, tracking of mathematics classes at Population A. By contrast, the analysis also showed more tracking of algebra at this grade level in the US than in any of the other participating countries.

The I-C chart displays a two-way ordering of OTL data: test items (rows) by classrooms (columns). The ordering is from top to bottom and from left to right in terms of item coverage. Here there are 30 algebra items reported upon by teachers in slightly more than 200 classrooms. Hence we have a display of ‘teacher coverage’ of the algebra items on the SIMS test. At upper left are classrooms with high teacher coverage. At lower right are classrooms with low teacher coverage.

I-C charts provide rich descriptions of the ‘goodness of fit’ of the SIMS test across countries. Not surprisingly, the charts for geometry are especially illuminating as the cross-country variations in content are displayed in this way. A key finding of SIMS was that cross-national differences in student achievement were found to be associated (though not perfectly) with Opportunity to Learn (Garden, 1987; Suter, 2000).

Within the context of SIMS, Richard Wolfe (1983) also devised a statistical software system in which data files with a hierarchy of observations can be stored with appropriate linkage of data with documentation.



*Volume II: Contexts and Outcomes of School Mathematics  
(Robitaille and Garden, eds., 1989)*

Enormous demands of data gathering were ever present as Volume II was in preparation--the New Zealand Centre bore the brunt of these burdens. And keep in mind that this was the 'pre-internet' era. One could only imagine, in fleeting flights of fantasy, the possibility of transmitting large data sets by telephone lines -- let alone by satellite signals!

After the data were collected and documented in the National Centers, they were sent to New Zealand for international processing. I vividly recall a meeting when Roy Phillipps was 'taking the roll' -- documenting the status of data submitted, center by center. At one point, a researcher reported: *'We have sent our data'*. Roy replied: *'Well, we have not yet received it.'* To which there was the response: *'We sent it by sea....Roy, Do you realize how expensive it would have been to send it by air?'*

Ah, the joys of international pre-internet project coordination!

SIMS provided a unique opportunity to look at changes that had taken place since the First IEA Study, reflecting, it is assumed, effects of curricular reform. For example, at the Population A level, emphasis on arithmetic had declined, while algebra and geometry had increased (Robitaille & Garden, 1989, p. 234).

At the Population B level (end of secondary school) the report notes wide differences between systems in the proportion of students who completed requirements needed for further study in post-secondary education. As the authors observe:

Such wide differences have important implications for curriculum, since a course which might be appropriate for the top 5 or 10 percent is unlikely to be appropriate for 30 percent of the age cohort, in the case of Canada (British Columbia); or for 50 percent of the age cohort, in the case of Hungary. (pp. 234-235)

### ***Volume III: Student Growth and Classroom Processes (Burstein, 1992)***

The third and final international report deals with the SIMS option known as the 'longitudinal study'. This version of the project, chosen by a subset of eight countries, featured three additional components to the standard, 'cross-sectional' IEA model. These methodological enhancements entailed: (i) providing a pre-test; (ii) drawing samples of at least two intact classrooms per school; and (iii) devising detailed, content-specific 'classroom process' questionnaires to be completed by the teacher.

Even though the longitudinal version of SIMS required a far greater data collection burden, we believed that the resulting added yield from this option would compensate for the increased effort. Following is a selective summary of those benefits.

#### **i. The pre-test**

A pre-test made it possible to measure student growth in achievement during the school year. As a result, the focus of SIMS was on growth as well as 'end of year status' (international ranking).

#### **ii. Sampling at least two classrooms per school**

With at least two classrooms per school, we were able to use variance decomposition (of the pre-test scores) to identify sources of variation in the achievement data. Specifically, if tracking is used within a school to assign some students to courses with less demanding subject matter content, this would be marked by greater within-school, between-class variation than in those schools where there is no tracking.

A key finding of the longitudinal study was that the country with the largest between-classroom component of variance was the United States, which tracked proportionately more students in mathematics than did the other participating countries. By contrast, Japan's variation arose almost entirely from between students within schools, indicating they do not track. (Recall the finding provided by the above-referenced I-C chart for Algebra in Japan. )

### iii. Classroom process questionnaires

We created detailed, content-specific classroom questionnaires for the teacher in order to provide information on the instruction that took place as the various topics were taught. The purpose was to look for association of the classroom process data with the accompanying growth measures made possible by the pre-test. As reported in Volume III there is great variety in instructional practice both within and between systems.

In summarizing SIMS accomplishments, Kifer (2011) has remarked:

As international studies of mathematics become less research and more indicator oriented, results of SIMS would caution those who place an inordinate emphasis on rank orders. SIMS documented diversity in what systems teach and when they teach it. SIMS demonstrated that growth and status are very different kinds of measures with very different correlates. It showed that teachers within a system respond in quite varied ways to questions about how they approach instruction. Perhaps the enduring results of SIMS are descriptions of a marvelously varied world of mathematics teaching and learning. (personal communication)

## Summary Reflections

What did SIMS add up to? It is likely that this question would be answered differently by each of us. For example, in the larger scheme of things, I am reminded of the influential Coleman Report (1966) that found that student background and socioeconomic status are more important in determining educational outcomes than is the quality of schooling itself (see, for example, Kiviat,2000). In contrast, SIMS clearly documents that curriculum and instruction do in fact make a difference in student achievement.

Exploring varieties of implications of SIMS, Westbury gathered together an impressive roster of scholars from such fields of specialization as

curriculum, evaluation, policy analysis and statistics, and produced the book entitled, *In search of more effective mathematics education: Examining data from the IEA Second International Mathematics Study* (Westbury, et al., 1994).

At the US National Center at the University of Illinois, and doubtless at others, visiting scholars came to spend time pursuing their own research interests in the improvement of teaching and learning mathematics in the light of SIMS findings.

At the national level, each IEA national center responded in its own way, of course. There was a myriad of national reports, distilling the findings and addressing national needs in the improvement of the teaching and learning of mathematics. Selected illustrative national reports include those for Japan (Sawada, 1981); Thailand (Thai National Committee, 1986) and New Zealand (New Zealand Department of Education, 1987).

In my role as Director of the US National Center, I am grateful for the team of mathematics educators who produced *The Underachieving Curriculum: Assessing US school mathematics from an international perspective* (McKnight, et al., 1987). From the beginning of SIMS, our national committee pledged that the SIMS findings would be targeted at the improvement of classroom instruction. The fact that our committee included F. Joe Crosswhite and John A. Dossey who, respectively, served successively as president of the National Council of Teachers of mathematics (NCTM) greatly reinforced this determination, to be sure. Curtis C. McKnight, who was National Research Coordinator, and is now Professor of Mathematics, University of Oklahoma, capably served as senior author of the US National Report. His well-crafted language was highly effective in translating SIMS findings into recommendations for improved classroom instruction.

Dossey (2011) has noted that SIMS, together with the data of the 1986 National Assessment of Educational Progress, portrayed a picture of school mathematics that was mired in drill and vocabulary, devoid of

understanding and applications. He goes on to remark:

Students were able to identify examples, but unable to apply generalizations. They were able to perform direct calculations, but unable to solve problems of medium complexity or check results of easier problems for reasonableness. (Personal communication)

Concerning instructional practice, the especially devised SIMS Classroom Processes surveys indicated that the dominant mode of teaching was one of turning text pages in a rush to 'cover material' with consequent diminished understanding. The patterns, problems, and principles of mathematics and its applications were lacking. Dossey concludes:

It was this flatscape that provided the base arguments for mounting the drive to produce the **NCTM Curriculum and Evaluation Standards for School Mathematics** (1989). Developed over the years between the initial availability of the SIMS data and the publication of *The Underachieving Curriculum*, the **NCTM Standards** document was the first national attempt at stating what all U S students should know and be able to do as a result of their K-12 education in mathematics...

Even a cursory glance at the NAEP results that followed the release of the **SIMS** report, which findings were bolstered by the influence of the **Standards** on school and state curricula, clearly reveals the impact of SIMS on school mathematics in the United States. (Dossey, Personal communication)

*The Underachieving Curriculum* report was acted upon by the profession in various ways--two are especially noteworthy. In one state, the supervisor of mathematics took special notice of the SIMS finding that because of extensive tracking, many eighth-grade students had only limited opportunities to learn algebra. Consequently, he was able to obtain funding from the National Science Foundation to replicate SIMS at the state level. A major purpose of his project was to provide data to accompany professional development activities that focused on expanding opportunities to learn algebra in the middle grades.

In another state, a leading educator sent a copy of the US National

Report to each school district with the intention that key individuals would give serious consideration to the report's many recommendations for improving school mathematics in their schools.

At a personal level, I reflect upon the opportunities that are made available to each person privileged to be a part of an IEA research project. The resulting circle of colleagues that are mine, both here in the U S and around the world have contributed immeasurably to my own professional development. For that I will be forever grateful.

## Epilogue

Let me make a personal note about Torsten Husén. Through the duration of several years of SIMS, Torsten Husén was a continued source of encouragement to me. On many occasions at national/international meetings, he would take me aside and ask how SIMS was doing. And at the conclusion of SIMS, he graciously made it possible for me to spend a two-week study leave in Stockholm with the understanding that I visit occasionally with various officials at the Swedish Ministry of Education and share various findings from the study. Fast forward: Recently, when IEA Executive Director Hans Wagemaker visited our campus, I was able to get caught up on current activities in IEA and was thereby prompted to send Professor Husén a note of appreciation for his importance to me personally as a mentor and role model. I did send such a message, hoping that Professor Husén was able to receive it. Eventually I did get a response, but very sadly, it was from his daughter. She let me know of Torsten's passing, but assured me that my e-mail had arrived in time.

## Endnotes

- i. Other persons directly associated with SIMS have now passed away: Sven Hilding, Gerry Pollock, Tamás Varga, Hans-Georg Steiner and Leigh Burstein.

- ii. I am grateful to John Dossey, Bob Garden, Skip Kifer, Curtis McKnight, Larry Suter, Ian Westbury and Richard Wolfe for assistance in preparing this chapter. I also thank the external reviewers for their help.
- iii. Technical support for preparing this chapter was provided by the University of Illinois Office for Mathematics, Science and Technical Education (MSTE), George Reese, Director. The MSTE office is a division of the College of Education, Mary Kalantzis, Dean.
- iv. Chapter photography by K. Sarah Spaulding, SIMS Illinois Administrative Assistant.  
Travers bio photograph by Elizabeth Kirchgessner, MSTE.

## References

- Burstein, L. (Ed.). (1992). *The IEA Study of Mathematics III: Student Growth and Classroom Processes*. London: Pergamon Press.
- Coleman, J. S. et al. (1966). *Equality of Educational Opportunity*. Washington DC: U. S. Department of Health, Education and Welfare, Office of Education.
- Freudenthal, H. (1975). Pupils' achievement internationally compared – The IEA. *Educational Studies in Mathematics*, 6, 127-186.
- Garden, R. (1987). Second IEA mathematics study. *Comparative Education Review*, 31, 47-68.
- Harnisch, D. L & Linn, R. L. (1981). Analysis of item response patterns: Questionable test data and dissimilar curriculum practice. *Journal of Educational Measurement*, 18 (3), 133-46.
- Husén, T. (1967). *International Study of Achievement in Mathematics: A Comparison of Twelve systems, Volumes I and II*. Stockholm: Almqvist and Wiksell.
- Kiviat, B. J. (2000). *The Social Side of Schooling*. *John Hopkins Magazine*. Retrieved from: <http://www.jhu.edu/jhumag/0400web/18.html>

- McKnight, C. C., Crosswhite, F. J., Dossey, J. A., Kifer, E., Swofford, J. O., Travers, K. J., & Cooney, T. J. (1987). *The Underachieving Curriculum: Assessing U. S. School Mathematics from an International Perspective*. Champaign, IL: Stipes.
- National Commission on Excellence in Education. (1983). *A Nation at Risk*. Washington, DC. Retrieved from <http://www2.ed.gov/pubs/NatAtRisk/index.html>
- National Council of Teachers of Mathematics. (1989). *Curriculum and Evaluation Standards for School Mathematics*. Washington, DC: Author.
- New Zealand Department of Education. (1987). *Mathematics Achievement in New Zealand Secondary Schools*. Wellington, New Zealand: Author.
- Robitaille, D. F. & Garden, R. A. (Eds.). (1989). *The IEA Study of Mathematics II: Contexts and Outcomes of School Mathematics*. London: Pergamon Press.
- Sawada, T. (Ed.). (1981). *Mathematics Achievement of Secondary School Students*. National Institute for Educational Research, Tokyo. (in Japanese)
- Servais, W. (1975). Continental traditions and reforms. *International Journal of Mathematics Education, Science and Technology*, 6 (1), 37-58.
- Suter, L. (2000). Is Student Achievement Immutable? Evidence from International Studies on Schooling and Student Achievement. *Review of Educational Research (RER)*, 70, 529-545.
- Thai National Committee for the Second International Mathematics Study. (1986). *The Evaluation of Mathematics Teaching and Learning in Mattayan Suksa 2 Classes*. The Institute for the Promotion of Teaching Science and Technology, Bangkok. (In Thai).
- Travers, K. J. (1983). A Comparative Review of Curriculum: Mathematics and International Studies in the Secondary Schools of Five Countries. Appendix C: Commissioned papers. *A Nation at Risk*. Washington, DC. Retrieved from <http://www2.ed.gov/pubs/NatAtRisk/index.html>



- Travers, K. J. & Westbury, I. D. (Eds.). (1989) *The IEA Study of Mathematics I: International Analysis of Mathematics Curricula*. London: Pergamon Press.
- Westbury, I., D. et al. (Eds.). (1994). *In Search of More Effective Mathematics Education: Examining data from the IEA Second International Mathematics Study*. Norwood, New Jersey: Ablex.
- Wilson, J. W. (1976). *A second international study of achievement in mathematics*. Discussion paper prepared for Basic Skills Group, National Institute of Education. (mimeo)
- Westbury, I. D. (Ed.). (1980). *Comparative Studies of Mathematics Curricula: Change and Stability, 1960-1980*. Proceedings of a conference jointly organized by the Institute for the Didactics of Mathematics (IDM) and the International Mathematics Committee of the Second International Mathematics Study of the International Association for the Evaluation of Educational Achievement (IEA), January 7-11, 1980. Bielefeld, FRG: Institut für Didaktik der Mathematik der Universität Bielefeld.
- Wolfe, R. (1983). *Integrating data and documentation in a multi-national research project: The IEA second international mathematics study*. Retrieved from: [http://portal.acm.org/citation.cfm?id=1116776&CFID=12630893&CF\\_TOKEN=83128108#citedby](http://portal.acm.org/citation.cfm?id=1116776&CFID=12630893&CF_TOKEN=83128108#citedby)

## CHAPTER 4

# **The IEA Studies in Civic Education: Their History from the Late 1960s to 2005**

*Judith Torney-Purta & John Schwille*

### **The Early Days of IEA and its Civic Education Study (late 1960s to mid 1980s)**

The co-authors of this contribution about the history of IEA took different routes into leadership in the civic education studies. Judith began participating in the IEA Civic Education Study that was part of the Six Subject Area Survey in the late 1960s. She had recently completed a book and a doctoral dissertation dealing with children's political attitudes at the University of Chicago and was an Associate Professor of Psychology at the University of Illinois in Chicago. Starting as a Steering Committee member, she wrote some of the cognitive items used in the test and made major contributions to designing attitudinal items (Anti-Authoritarianism, Support for Women's Rights, Open Classroom Climate for example). There was also a new instrument (initiated by Bram Oppenheim of the London School of Economics) called "How Society Work," in which students were asked to rate the functioning of institutions such as elections, the democratic system of government, and the police. In the early 1970s Judith spent part of two summers and one December at IEA Headquarters in Stockholm first designing the analysis and then writing the book, *Civic Education in Ten Countries: An Empirical Study* (Torney, Oppenheim & Farnen, 1975). Remember that this was a time when doing computer analysis meant physically transporting data and outputs back and forth to a computer center and when typewriters were the unwieldy processors of words.

John began his work at IEA with a Spencer Fellowship in 1972 to ana-

lyze the Six-Subject Survey data just at the point when the first international reports were being prepared at the IEA secretariat in Stockholm. He had been intrigued by IEA since his graduate course in comparative education at the University of Chicago in 1963 when Professor Arnold Anderson announced that at long last comparative education was going to have rigorous dependent measures of educational outcomes in the First IEA Mathematics Study. So when John saw an advertisement for the Spencer Fellowships in 1972, he was eager to apply. In his first conversation with Judith she noted that he did not have experience in the intersection of psychology, political science and education that was needed. But as soon as she read his excellent proposal, she changed her mind, beginning nearly 40 years of collaboration and friendship. In Stockholm, he contributed to all of the six subject surveys as a Senior Research Officer and he was asked to stay on to head the team working on a Six-Subject Survey databank to be used for secondary analysis, but his special interest was in civic education. He served as an intermediary to the brilliant but independent minded and somewhat undisciplined team of programmers who were doing the analyses for the civics volume to be published in 1975. He also authored the chapter dealing with the regression analysis of the civics cognitive test outcomes. This was an opportunity to use a technique developed by Newton and Spurrell to moderate the standard IEA analyses, which in Judith and John's view, did not allow full exploration of the characteristics of schooling in explaining differences in scores on the tests developed for the Six-Subject Survey.

Our entry in the recently issued *International Encyclopedia of Education* reflects on the atmosphere of the 1970s in IEA when this first civic education study was conducted:

It is difficult to recapture today the concerns that surrounded this domain in the midst of the Cold War. What counted as legitimate civic education in one country was not what counted in countries with different ideologies. Measurement was daunting in the civic education domain, where attitudes were

important as desired outcomes of civic education, and where model standards for measurement of knowledge were rare. In other words, this study was a bold move with risks both for the researchers and for IEA as an organization (Torney-Purta, Amadeo, & Schwille, 2010, p.656).

In the period from the late 1970s until the late 1980s the topic of civic education was not central to IEA's interests. A few articles were published using these data, for example comparing knowledge of international topics with knowledge of national topics in several countries (Torney, 1977). This was intended to contribute to the growing debate about global education in the U.S., since unlike the other Western democracies (who performed about equally well on items with national topics and those with international topics), U.S. students were better at answering national items and also were much more interested in national news than in international news. The fact that endorsement of democratic values was high in West Germany and the Netherlands, but that students in the United States had relatively low scores on these scales (including support for women's rights) never received the attention that we thought it deserved.

During the period when the U.S. panel Nation at Risk was convened, we collaborated on an analysis resulting in a report to the Commission and an article entitled "Civic values learned in school: Policy and practice in industrialized countries." (Torney-Purta & Schwille, 1986). During this period IEA was making its reputation with studies of mathematics, science, and reading literacy. John, then at Michigan State, was working on the second IEA math study. Judith, then at the University of Maryland, was using cognitive psychology approaches in small scale studies of young people's political cognition in the U.S. (Torney-Purta, 1989). However, both retained an interest in how large scale international studies might play a role in educational policy (Torney-Purta, 1990). Judith also participated in a planning committee for an IEA Values Education Study, which never took place but which had a 2-phased design that she later remembered.

## **IEA and Civic Education (CIVED) in the Post Cold War Era (the late 1980s to 2008)<sup>1</sup>**

There is no date that is a clear inflection point, though several things happened in the late 1980s. We met individuals who would later become central in civic education studies. While Rainer Lehmann (whose background was in history education) was attending an IEA Reading Literacy meeting in Washington, D.C. he asked to observe one of Judith's studies of political cognition. Soon after TIMSS was conceived (in 1988), the National Academy of Sciences in the U.S. established the Board on International Comparative Studies in Education (BICSE) to oversee U.S. participation. Both of us accepted invitations to become members. What we learned from those experiences later helped us establish a strong foundation for the CIVED Study as well as reinforcing our continuing interest in policy implications of all the IEA studies and enlisting us in rejoinders to defend comparative studies against what we believed to be unwarranted criticism (Bradburn, Haertel, Schwille, & Torney-Purta, 1991). The software developed by Andreas Schleicher for the Reading Literacy Study made IEA data processing and management more efficient than before. The methodological standards established by IEA for the TIMSS and PIRLS studies during this period were to have long-term implications.

In this period civic education was having a kind of renaissance. During the late 1980s the National Assessment of Educational Progress (NAEP) in the United States tested the civic knowledge of representative samples of 4th, 8th, and 12th graders. Niemi and Junn, who were political scientists, reanalyzed the 1988 NAEP data and suggested that current education was inadequate in this area (Niemi & Junn, 1998).

---

1. In what follows we are focusing primarily on an informal relating of the history of the project; publications in the references can be consulted for more substantive material.

Similar discontents with the effectiveness of civic education and the level of youth participation were voiced in England and Australia in this period.

The most influential political event in this period was the fall of the Berlin Wall and the collapse of Communism across Central and Eastern Europe, however. Questions were raised about the extent to which the educational systems of these countries were prepared to teach democracy to young people. Would it be possible to slide Marxism out of the curriculum and replace it with democratic theory? What about teachers who had grown up teaching with one set of methods and would be asked to teach in a new way and covering material with which they had limited familiarity? What about young people who had been taught in their families to voice their views about social and political issues only in whispers? The Council of Europe and other organizations began to hold meetings bringing together Eastern European and Western European educators. At those meetings Judith met Gita Steiner-Khamsi (then a lecturer in Switzerland) and Barbara Malak-Minkiewicz (living in the Netherlands after being the Press Spokesperson for Solidarity in Poland). Neither had ever heard of IEA, though they were intrigued by its potential when it was described .

Judith, because she had been the leader of the previous study, was invited in 1993 to prepare a proposal for a civic education study to present to the IEA General Assembly meeting in Spain. Such a proposal had been suggested by some of the Eastern European IEA delegates whose countries had participated in math or reading studies. They saw the relevance of the comparative methodology for a topic high on their agendas, that was the need to know how the next generation of citizens could be prepared for democracy. Declining levels of political interest and participation among young people motivated some delegates from Western Europe to support a civic education study. In advance of the meeting, Judith had prepared a proposal for a relatively traditional IEA test and survey combination. The day before

the presentation when she looked around the room at the GA delegates she realized how little was known about civic education aims or practices in post-Communist countries. She handwrote a new transparency to suggest a two-phased study, in which the first phase would be a set of national cases studies (a qualitative approach). She argued that this would allow the planners to learn about the substantive issues in different countries (leading to better test and survey items in the second quantitative phase). The General Assembly agreed to go ahead with the planning on this basis. Immediately after the vote Rainer Lehmann, who was representing Germany, approached Judith to offer his collaboration in this effort.

She invited John to join her efforts to formulate guidelines for these studies. These guidelines were based on the idea that conceptions of civic education could vary much more across countries and even within countries than would be the case in subjects like mathematics and science. To capture these differences and make it possible to take them into account in designing and developing a test and survey for a second phase to the study, a coordinated case study design was used. Each participating country team was asked to write a case study for the country which would respond to a list of general and specific questions about the nature of civic education in the country. The drafts of these cases studies were then reviewed and given feedback by members of the international steering committee, primarily Gita Steiner-Khamsi and Barbara Malak. We expected these case studies to be so diverse that it would be hard to agree on the domains that could be addressed across countries in this second civic education study, but this turned out not to be the case.

Developing international scientific projects by consensus between project team members is something often sought but difficult to achieve as documented by the Committee on Psychological Science at the National Academy of Sciences in the U.S., which recently issued a report entitled *International Collaborations in the Behavioral and Social Sciences* (2008).

At the first meeting of the CIVED National Research Coordinators in the Netherlands in 1994 it became clear that a collaborative atmosphere would be challenging to establish and have to be carefully nurtured. Some of the Eastern Europeans were skeptical of the motives of the organizers of the project. They feared that the Western Europeans and the U.S. simply wanted to show what poor prospects existed for democracy in Eastern European countries. John remembers being upbraided by one of the authors of a national case study for presuming that he knew enough about the country in question to review the materials. The confrontational tone of a speech to the group from an interim Executive Director of IEA didn't help the meeting's atmosphere.

It is impossible to say even in hindsight what turned the NRC meeting and then the project from an adversarial atmosphere to a positive collaboration. It may have been a focus group led by Barbara Fratzak-Rudnicka, a Polish sociologist, where people were asked about their first political memories, allowing many commonalities and quite a few emotional moments to be shared. Another contribution was Barbara Malak-Minkiewicz's masterful lecture on interviewing techniques needed in Phase 1. That these two major figures from a post-Communist country saw the potential value of the project contributed to a more positive view among many delegates. Isabel Menezes, from Portugal, proposed a graphic model for the project, which was amended and then accepted as the Octagon Model (guiding the project and appearing in nearly every publication). In addition, rather than being asked to accept a pre-established content framework, all those in attendance were asked to vote on a series of topics that might be covered in a test and survey. There was more consensus than many had expected, and the group agreed to take steps forward in planning (and IEA supported these financially and organizationally starting in 1994).

The Planning Committee (later called the Steering Committee) was chaired by Judith Torney-Purta (U.S.) with Rainer Lehmann (Germany) as the International Coordinator. Committee members included the following: Barbara Fratzak-Rudnicka (Poland); Georgia



Kontogiannopoulou-Polydorides (Greece); Bruno Losito (Italy); Barbara Malak-Minkiewicz, (by this time a member of the IEA Secretariat in the Netherlands); Ingrid Munck (Sweden); Hans Oswald (Germany); John Schwille (U.S.); Gita Steiner-Khamsi (Switzerland and U.S.); Lee Wing On, (Hong Kong SAR). In addition, Heinrich Mintrop served as a consultant for the teacher questionnaire; Ray Adams was Ex-Officio from the IEA Technical Executive Group; Wolfram Schulz and Jo-Ann Amadeo served as Associate Coordinators. The committee met before and after the National Research Coordinators' meetings to plan and then to implement decisions. Their most crucial roles were in deciding what was to be included in the instruments and in planning the international analyses and reports.

The first phase, taking place over about three years, also allowed each national research coordinator to have a chapter publication from their national case study to show for their early efforts. The Pew Charitable Trust partially funded this phase. Jo-Ann Amadeo, a colleague from the University of Maryland, became a valued collaborator. The nickname for the book is "the brick," since it is 627 pages long (Torney-Purta, Schwille & Amadeo, 1999). In addition, Gita Steiner-Khamsi prepared an edited book of reflections across case studies (Steiner-Khamsi, Torney-Purta & Schwille, 2002) with a chapter on the elusive dimensions of civic education in schools (Schwille & Amadeo, 2002).

"The brick" had an unexpected success. It was named an Outstanding Academic Title of 2000 by the journal *Choice*, which is used by bibliographers in U.S. research libraries to select books. In the year 2000, about 600 books received this award from the 6700 titles reviewed and the 25,000 volumes submitted to this journal. To our knowledge, this is the only major academic book award received by an IEA publication.

The case study materials that were being submitted from participating countries were the major source for the development of the conceptual framework on which the test and survey were constructed. There were three parts to this framework: I: Democracy and its definition;

democratic institutions; citizenship rights and duties; II: National identity; international relations; III: Social cohesion and diversity. In order to keep the process on track with countries' needs, each section included direct quotations from the extensive case study materials that served to clarify and illustrate content covered by these three domains

The responsibility for item preparation was distributed. There were inadequate funds to pay item writers, so all the members of what was then called the Planning Committee were asked to work on them and were instructed to pay close attention to quoted material in the conceptual framework document. Judith's cognitive research provided some examples of young people's misconceptions about social and political issues. We collected items from Bruno Losito in Italy and revised them for international respondents. John constructed trial items to fit aspects of the framework that called for new ways of measurement, especially the more conceptual parts of the framework. Lee Wing On kept reminding us that sometimes Asian democracy provided a unique context. Georgia Polydorides and Gita Steiner-Khamsi suggested the use of political cartoons and newspaper texts as inspirations for measures of interpretive civic skills; at a meeting in Germany Rainer Lehmann provided stacks of newspapers to stimulate construction of these items (and also helped us think through issues with items about distinguishing facts from opinions). Barbara Fraczak-Rudnicka wrote several items measuring civic skills, including using an election leaflet as the material on which questions were based. We had to check the content of every item so that it would not be biased toward the country where it had been developed (essential in this subject area where political institutions differ so much). And Ingrid Munck reminded us of the analytic issues to keep in mind.

With funding from a German foundation, (the DFG) the International Coordinating Center was established at the Humboldt University of Berlin under Rainer Lehmann's leadership. Wolfram Schulz joined the project as associate coordinator a couple of years later. We succeeded (in spite of relatively low levels of funding compared with other proj-

ects) to meet IEA standards for the instrument. More than two dozen countries participated (11 post-Communist countries, 12 other European countries, 2 Latin American countries, Australia, Hong Kong SAR, and the United States).<sup>2</sup> Approximately 90,000 fourteen-year-old students from 28 countries were administered a test and a survey, each taking about 40 minutes. These data were collected in 1999, findings were released in 2001, and reported in the volume entitled *Citizenship and Education in Twenty-eight Countries: Civic Knowledge and Engagement at Age Fourteen* (Torney-Purta, Lehmann, Oswald, & Schulz, 2001). School heads and teachers had also provided information, covered in a chapter by Losito and Mintrop. In 2000, over 50,000 upper secondary school students from 16 countries received a similar test of civic knowledge and skills and the same survey of civic attitudes and behaviors; these findings are reported in *Civic Knowledge and Engagement: An IEA Study of Upper Secondary Students* (Amadeo, Torney-Purta, Lehmann, Husfeldt, & Nikolova, 2002). Both reports are at <http://www.terpconnect.umd.edu/~iea> .

Both groups of students were tested on their civic knowledge and skills. They also were surveyed on their concepts of democracy and government; their attitudes related to trust in institutions, political efficacy, and opportunities for immigrants and women; and their expected participatory actions relating to politics. A final part of the student survey assessed the students' perceptions of the climate for discussion in their classrooms as well as other background variables. In addition, an internationally relevant list of organizations was developed and students were asked to indicate to which groups they belonged (Torney-Purta, Amadeo, & Schwille, 2010, p 658).<sup>3</sup>

---

2. Australia, Belgium (French), Bulgaria, Chile, Colombia, Cyprus, Czech Republic, Denmark, England, Estonia, Finland, Germany, Greece, Hong Kong (SAR), Hungary, Italy, Latvia, Lithuania, Norway, Poland, Portugal, Romania, Russian Federation, Slovak Republic, Slovenia, Sweden, Switzerland, and the United States.

3. This source contains a summary of the findings and their role in IEA studies.

Funding from the William T. Grant Foundation to the Department of Human Development at Maryland helped a great deal. Lonnie Sherrod, Vice President of the foundation at the time, believed that civic engagement was an under-studied area that fit well into the research priorities. After leaving the Grant Foundation he became the senior editor of a broadly based handbook on civic engagement research (Sherrod, Torney-Purta, & Flanagan, 2010). The CIVED study receives considerable attention in this volume: cross-national views of political participation, immigrants' participation, the role of classroom discussion, the use of qualitative and quantitative research methods, and the developmental theory basis of the CIVED Study.

With W.T. Grant funds we were able to prepare an Executive Summary for report releases in 2001 at the Comparative and International Education Society meetings in Washington and in 2002 at the International Society of Political Psychology in Berlin. Several publications summarized these results for different audiences (Torney-Purta, 2001; Torney-Purta, 2002, Torney-Purta, Barber & Richardson, 2004). We have relied very little on country rankings and have had a focus on analysis that can help illuminate students' learning processes. We have used HLM to look at Latino students in the U.S. (Torney-Purta, Barber & Wilkenfeld, 2007), students' volunteering in 4 countries (Torney-Purta, Amadeo, & Richardson, 2007), support for human rights in 27 countries (Torney-Purta, Wilkenfeld & Barber, 2008), gender differences in efficacy and support for women's rights (Barber & Torney-Purta, 2009), and support for immigrants' rights in 25 countries (manuscript in progress). We have used person-centered cluster analysis of attitudinal items in five Western European and five Eastern European countries (Torney-Purta, 2009). This approach has identified the following profiles; Social Justice Oriented Students, Conventionally Political Students, Indifferent Students, Disengaged Students and Alienated Students (characterized by strongly anti-immigrant and anti-minority attitudes).

In the United States during this period civic education generally took

on a higher profile resulting in the establishment of the Center for Information on Civic Learning and Engagement (CIRCLE – <http://www.civicyouth.org>). This is currently located at Tufts University in Massachusetts. Among the publications found there are analyses linking teachers with students (Torney-Purta, Barber, & Richardson, 2005), and a publication about 21st Century Workplace Skills (Torney-Purta & Wilkenfeld, 2009). Judith and Jo-Ann Amadeo coordinated a workshop on data analysis using CIVED attitudinal data for about a dozen researchers from low-income countries in conjunction with the International Psychology Congress in Berlin in 2008.

Several of the dissemination efforts have been aimed at audiences' interested in policy; an especially good example is a specialized analysis of the IEA data from Chile, Colombia, Portugal, and the United States published in *Strengthening Democracy in the Americas through Civic Education: An Empirical Analysis of the Views of Students and Teachers* (2004). The analysis, supported by the Organization of American States, identified problems in this region with students' lack of awareness of threats to democracy and poor understanding of citizens' rights that had been obscured when all 28 countries were analyzed. The Chilean National Commission on the Reform of Citizenship used this analysis in arguing for curricular reform. See also results from the Pacific Rim countries that participated in CIVED (Kennedy, Hahn, & Lee, 2008). A special issue of the British journal *Citizenship Teaching and Learning*, considered Latin America (Reimers, 2007) and the post-Communist countries (Malak-Minkiewicz, 2007) among others.

Special efforts have been made to support secondary analysis by a wide range of researchers. CEDARS (Civic Education Data and Researcher Services) was established at the University of Maryland to provide assistance to researchers, graduate students, and policy-makers' staff; additional IRT scales were formulated and added to the dataset (Husfeldt, Barber & Torney-Purta, 2005). These efforts culminated in the acceptance of the data for systematic dissemination by the ICPSR at the University of Michigan (<http://dx.doi.org/10.3886/ICPSR21661>).

Our consistent efforts to make the project internationally collaborative were generally successful. In fact when the National Academy of Sciences in the United States issued a report analyzing successful international collaborations, its authors who came from several social and behavioral sciences proposed that international projects should be constructed with three phases (National Academy of Sciences, 2008). The first phase requires trust and consensus building among participants, identifying assumptions, and identifying both commonalities and differences in approaches. This is clearly what happened in the first phase of the CIVED study. In the second phase, according to the Academy Report, the research plan is carried out. This corresponds to the CIVED test and survey, and was certainly facilitated by existing IEA structures and networks of specialists. During the third phase, according to the report, the focus should be on disseminating and publishing findings and on releasing a documented data set to scholars. IEA studies, including CIVED, have set valuable precedents in the release of instruments and data and the encouragement of collaboration in publications.

### Summary

Meeting the challenges of the first two civic education studies was somewhat different from working on the better known IEA studies in mathematics, science and literacy. Our work was initiated just as the protests involving many students in the 1960s were receding and continued through intermittent wars, economic crises, breakthroughs in attitudes about women in politics, substantial changes in the political climate and regimes of participating countries, tendencies toward greater globalization and the breakup of the Soviet Union. Then we entered into an era in which the ideologies of markets and multiparty democracies were at least temporarily dominant across the world.

At the same time, the capacity to do large-scale social and educational research increased dramatically through more advanced methods of

analysis, progress in computing and ease/cost of communication, but most of all in the growing level of competence among researchers throughout the world. This was growth in human potential for which IEA itself could take some credit. At the place where this account ends, we can say that comparative research on civic education had established itself as a major and continuing area of interest to policy-makers, researchers and educators alike. All are looking for more precise and valid ways of understanding the conceptions, attitudes, knowledge and actions of young people on whose learning our civic futures depend.

In conclusion, we are energized by the excitement and enthusiasm of our longtime Swedish collaborator Ingrid Munck who often says that she truly believes that we are making a contribution to a better future through new understanding about what democracy means to young people and how they learn to be citizens. This enthusiasm is shared by many of us even though we know how difficult it is to do high quality research in this area and to ensure that this research has an impact.

## References

- Amadeo, J., Torney-Purta, J., Lehmann, R., Husfeldt, V., & Nikolova, R. (2002). *Civic knowledge and engagement among upper secondary students in sixteen countries*. Amsterdam: International Association for the Evaluation of Educational Achievement.
- Barber, C. & Torney-Purta, J. (2009). Gender differences in political efficacy and attitudes to women's rights influenced by national and school contexts: Analysis for the IEA Civic Education Study. In D. Baker & A. Wiseman (Eds.), *Gender equality and education from international and comparative perspectives. International Perspectives on Education and Society, Vol 10*, (pp. 357-394), Bingley, U.K: JAI/Emerald Group Publishing.
- Bradburn, N., Haertel, E., Schwille, J., & Torney-Purta, J. (1991). A rejoinder to "I never promised you first place." *Phi Delta Kappan*, 72 (10), 774-777.

- Husfeldt, V. , Barber, C., & Torney-Purta, J. (2005). Students' social attitudes and expected political participation: New scales in the enhanced database of the IEA Civic Education Study. College Park, MD : CEDARS. <http://www.terpconnect.umd.edu/~iea>
- Kennedy, K., Hahn, C., & Lee, W. (2008). Constructing citizenship: Comparing the views of students in Australia, Hong Kong and the United States. *Comparative Education Review*, 52 (1), 53-59.
- Malak-Minkiewicz, B. (2007). Civic education in times of change: The post-Communist countries. *Citizenship Teaching and Learning*, 3 (2), <http://www.citized.info>
- National Academy of Sciences (2008). *International collaborations in the behavioral and social sciences*. Washington, D.C.: National Academies Press.
- Reimers, F. (2007). The study of civic education when democracy is in flux; Analyzing the of empirical research in policy and practice in Latin America. *Citizenship Teaching and Learning*, 3 (2), 1-82.
- Schwille, J. & Amadeo, J. (2002). The paradoxical situation of civic education in schools: Ubiquitous and yet elusive. In Steiner-Khamsi, G., Torney-Purta, J., & Schwille, J. (Eds.). *New paradigms and recurring paradoxes in education for citizenship: An international comparison* (pp. 105-139). Amsterdam: Elsevier.
- Sherrod, L., Torney-Purta, J., & Flanagan, C. (2010). *Handbook of research on civic engagement in youth*. Hoboken, NJ: John Wiley.
- Steiner-Khamsi, G., Torney-Purta, J., & Schwille, J. (Eds.). (2002). *New paradigms and recurring paradoxes in education for citizenship*. Amsterdam: Elsevier Science (JAI Press).
- Torney, J., Oppenheim, A. N., & Farnen, R. (1975). *Civic education in ten countries: An empirical study*. New York: Halsted Press of John Wiley.
- Torney, J. (1977). The international knowledge and attitudes of adolescents in nine countries: The IEA civic education survey. *International Journal of Political Education*, 1, 3-20.



- Torney-Purta, J. (1989). Political cognition and its restructuring in young people. *Human Development*, 32, 14-23.
- Torney-Purta, J. (1990). International comparative research in education: Its role in educational improvement in the U.S. *Educational Researcher*, 19, 32-35.
- Torney-Purta, J. (2001). Civic knowledge, belief about democratic institutions, and civic engagement among 14-year-olds. *Prospects (A UNESCO Journal)*, 31, 279-292.
- Torney-Purta, J. (2002). The school's role in developing civic engagement: A study of adolescents in twenty-eight countries. *Applied Developmental Science*, 6, 202-211.
- Torney-Purta, J. (2009). International psychological research that matters for policy and practice. *American Psychologist*, 64, 822-837.
- Torney-Purta, J., & Amadeo, J. (2004). *Strengthening democracy in the Americas through civic education: An empirical analysis of the views of students and teachers*. Washington, D.C.: Organization of American States.
- Torney-Purta, J., Amadeo, J., & Richardson, W. (2007). Civic service among youth in Chile, Denmark, England and the United States: A psychological perspective. In M. Sherraden & A. McBride (Eds.), *Civics service worldwide: Impacts and inquiries*. Armork, NY: ME. Shape.
- Torney-Purta, J., Amadeo, J., & Schwille, J. (2010). IEA study in civic education. In P. Peterson, E. Baker & B. McGaw, (Eds.), *International encyclopedia of education (3rd edition)*: Vol 4. (pp. 656-62). Oxford: Elsevier Limited.
- Torney-Purta, J., Barber, C., & Richardson, W. K. (2004). Trust in government related institutions and political engagement among adolescents in six countries. *Acta Politica*, 380-406.
- Torney-Purta, J., Barber, C., & Richardson, W. (2005). *How teachers' preparation relates to students' civic knowledge and engagement in the United States: Analysis from the IEA Civic Education Study*. College Park,

MD: Center for Information and Research on Civic Learning and Engagement.

- Torney-Purta, J., Barber, C. & Wilkenfeld, B. (2007). Latino adolescents' civic development in the United States: Research results from the IEA Civic Education Study. *Journal of Youth and Adolescence*, 36, 111-125. [Winner of the Richard M. Wolf Memorial Award from the International Association for the Evaluation of Educational Achievement honoring the best article using IEA data published in 2007].
- Torney-Purta, J, Lehmann, R., Oswald, H., & Schulz, W. (2001). *Citizenship and education in twenty-eight countries: Civic knowledge and engagement at age 14*. Amsterdam: International Association for the Evaluation of Educational Achievement.
- Torney-Purta, J., & Schwille, J. (1986). Civic values learned in school: Policy and practice in industrialized countries. *Comparative Education Review*, 30, 30-49.
- Torney-Purta, J., Schwille, J., & Amadeo, J. (Eds.). (1999). *Civic education across countries: Twenty-four case studies from the IEA Civic Education Project*. Amsterdam: International Association for the Evaluation of Educational Achievement.
- Torney-Purta, J., & Wilkenfeld, B. (2009). *Paths to 21st Century competencies through civic education classrooms*. Chicago: American Bar Association, Division on Public Education and Campaign for the Civic Mission of Schools. Retrieved from <http://www.civicyouth.org>
- Torney-Purta, J., Wilkenfeld, B., & Barber, C. (2008). How adolescents in twenty-seven countries understand, support and practice international human rights. *Journal of Social Issues*, 4, 857-88.



## CHAPTER 5

# The International Study of Writing

*Sauli Takala*

### Introduction

I have been asked to write an account of the IEA International Study of Writing for the present volume. The untimely passing away of Alan C. Purves, who took the initiative to propose and plan the study, and of Anneli Vähäpassi, who chaired the International Steering Committee, meant that I stood next in line as I was the former International Project Coordinator. My chapter will basically be a narrative of the main activities, but will also provide some glimpses of events and processes which are not reported in the two published books on the subject (edited by Gorman, Purves & Degenhart, 1988; and by Purves, 1992). In addition, I will give some examples of the dissemination of information about the project and of the impact it seems to have had. This will be a personal account and it will be a selective one, as after almost 30 years, I had access to only a small part of relevant sources. Memory is also selective and this became very clear as, browsing the documents I did have access to, I began to recall events I had forgotten and I relived emotions I had felt (cf. Takala 1985/1986). For a perceptive discussion of the IEA written composition study, see also Hartmann (1995) and de Glopper (1995).

I will begin with a brief look at the activities of IEA in the area of humanities prior to the writing study. I recall that this prior history occasionally emerged in discussions within the writing project. It is also very clear that our study was planned in close adherence to the 'IEA model' in comparative studies of educational achievement, and so I will discuss the design and the conduct of the study, focusing on the various conceptual models we developed for the study. These are prob-

ably the study's most significant contribution to language education. I then discuss funding at some length, as problems related to resources led to delays, and a variety of problems and tensions emerged requiring many ad-hoc solutions. The main findings are briefly summarized and the chapter concludes with some reflections on the endeavor.

### **Context of the Study**

During the Six Subject Survey, IEA conducted pioneering studies of English and French as a foreign language. I was involved in the Six Subject Survey in Finland in several capacities: as a translator of many instruments; as a member of the national team for English, writing some items that were ultimately included in the English tests; as a tester who administered tests in some schools, carrying with me a tape recorder as not all schools had one to deliver the test of listening comprehension; and finally, as a co-author of the national report. The studies coordinated by John B. Carroll for French (1975) and Glyn Lewis and Carolyn Massad for English (1975) provided a valuable model on how to conduct a study of language achievement and how to relate the results to a number of contextual and individual variables. It is noteworthy that even writing and speaking were tested – as even today, a test of speaking is often excluded from large-scale surveys.

More immediately relevant for the writing project were the study of reading comprehension in the mother tongue (Thorndike, 1973) and the study of literature (Purves, 1973). The study of literature, in particular, was of interest as it highlighted the cultural embeddedness of literature instruction and of literary appreciation. Alan Purves had continued the exploration of the data and published a study of reading and literature education (Purves, 1979) in New Zealand.

An extensive plan to initiate a project on "Language of instruction and school achievement of linguistic minorities" was proposed by Dr Stacy Churchill, OISE, at the Enschede General Assembly in August 1983. The plan was not approved. In retrospect, this is regrettable as the

study would have anticipated very well the current, and increasing, concern with the topic. Another partly missed opportunity in the area of language education was the Language Education Study (LES, 1993-1996), which managed to implement only the first phase of the planned three phases. Phase 1 gathered information on language education at the national level (sociolinguistic context, language policy, language curriculum and assessment, language teaching, and professional support), school level (characteristics of schools and language teachers, provision and organization of language learning in schools, organization of curriculum, and assessment in the classroom), and student level (level of proficiency, attitudes, and aspirations). The interesting results were reported in a survey edited by Dickson and Cumming (1996). It is only now, almost 20 years later, that an international survey of achievement in foreign languages is being carried out as an EU project-- European Indicator of Language Competence (EILC). Information on this project is available at: [http://europa.eu/legislation\\_summaries/education\\_training\\_youth/lifelong\\_learning/c11083\\_en.htm](http://europa.eu/legislation_summaries/education_training_youth/lifelong_learning/c11083_en.htm)

My background is second/foreign language education, which has a very long history of international interaction and exchange. Since the 1960s, applied linguistics has provided many new insights but it also has tended to be dominated by experts in foreign language (L2) education. While the L2 profession has thus been fundamentally internationally minded, at the time when the writing study was launched, it seemed that mother tongue education (L1) was more closely linked to the agenda of nation building and/or the construction and maintenance of national identity and values. It would appear that in comparison to L2 education, mother tongue educators had relatively few international contacts. In fact, the International Mother Tongue Education Network (IMEN) was not set up until 1984. A good survey of its work can be found in Herrlitz, Ongstad & van den Ven (2007).

There were some exceptions to this pattern. In the summer of 1966 an Anglo-American seminar was arranged on the teaching of English.

Substantial differences were discovered between the attitudes towards English of teachers in England and America. Squire and Applebee (1969) published a survey of teaching of English in the United Kingdom and made a number of comparisons to their 'native' American practice. Note, however, that this rare instance of international comparison dealt with the same language, English, and was thus limited to a section of the Anglophone community.

This, in a nutshell, was the situation in the late 1970s when Alan Purves, who had a unique knowledge of the international context due to his many contacts and networks, started planning a truly international new project in the domain of the mother tongue. The fact that he chose to focus on writing was probably related to the fact that there had been a strong revival of interest in the assessment/testing of free writing to supplement - or replace - the indirect, 'objective' testing of writing, which had emerged especially in the United States as a response to the allegedly disastrously low reliability in rating free writing. Writing had also become an important area for explorations in cognitive psychology. In retrospect, it seems to me that there had been sufficient progress made both in the theory and practice of 'free' writing and in its assessment, and as a result the time was ripe for a comparative study of writing (e.g., Petrovsky & Bartholomae, 1986).

### **Project Approach**

Alan Purves formed an international team that drafted a proposal to be submitted to the General Assembly, held at the University of Jyväskylä, Finland in August 1980. The other team members consisted of Dr Eva Baker from UCLA, Mrs Anneli Vähäpassi from the University of Jyväskylä and Dr Hildo Wesdorp from SCO in the Netherlands. Eva Baker and her colleagues at the Center for the Study of Evaluation at UCLA had carried out extensive work on the assessment of writing (eg., Smith, 1978); Hildo Wesdorp was a recognized expert in mother tongue education in the Netherlands, who had been

working on a thorough review of methods used in the assessment of writing (Wesdorp, 1978); and Anneli Vähäpassi was a leading expert in Finland who had extensive experience conducting national assessments in the mother tongue.

Their proposal was approved. There was a growing awareness of the importance of literacy, and there was increasing worry over the alleged deterioration of students' writing ability--frequently seen as an indication of inadequacies in the school systems. Introducing students to written language (reading and writing), and thus promoting literacy, had traditionally been seen as one of the principal tasks of the school. There was also a discernible trend in that some countries (e.g., the UK, Australia, the Netherlands, Sweden/Lindell 1980, Finland) had begun or were beginning to assess systematically the effectiveness of teaching and learning writing. However, most of the countries that ultimately participated in the study had never carried out a large-scale empirical survey of writing in their school systems.

Initially following the traditional IEA approach, the main idea was to compare the level of achievement in writing; as the project developed, however, doubts soon emerged about the feasibility of strict comparisons between countries on an indicator for achievement. This meant that ultimately the study paid increasing attention to contextual and cultural variation. This was, indeed, one of the research tasks from the very beginning, but a strong hope for reliable comparisons lingered in some participating countries. The study was accordingly designed to accomplish the following research tasks:

- to contribute to the conceptualization of the domain of writing, and particularly the domain of school-based written composition.
- to develop an internationally appropriate set of writing tasks and a system for assessing composition that would be applicable across countries, school systems and languages.
- to describe recent developments and the current state of



instruction in written composition in the participating countries/school systems (using very extensive curriculum and teachers questionnaires).

- to identify factors that would explain differences and patterns in the performance of written composition and other outcomes, with particular attention to cultural background, curriculum, and teaching practices.

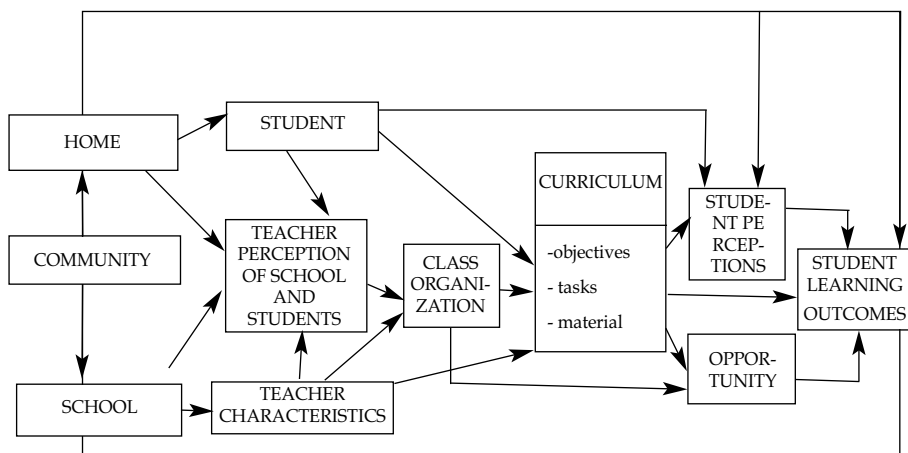
The study accordingly examined teaching and learning of written composition in the schools in order to identify the beliefs and conventions associated with written composition. The study also endeavored to find factors explaining differences and patterns in the performance of written composition and other outcomes, with particular attention on cultural background, curriculum, and teaching practices. Very extensive national, school, teacher and student questionnaires were employed to collect data for such explanatory analyses. Six types of writing were assessed (reflective, personal, philosophic, argumentative, persuasive, and literary) on four dimensions: style and tone, overall impression, content, and organization. The data were collected in 1985.

The design presented in Figure 1 shows that the writing study was planned using the traditional model employed in earlier IEA studies.

Participating countries included: Chile, England, Finland, Hamburg-Germany (FRG), Hungary, Indonesia, Italy, the Netherlands, New Zealand, Nigeria, Sweden, Thailand, the United States and Wales. The International Coordinating Center was initially located at the University of Illinois at Urbana-Champaign (UIUC).

The study included three populations: students near the end of primary schooling (Pop. A); students near the end of compulsory schooling (Pop. B); and students near the end of academic secondary school (Pop. C).

Participating countries selected writing samples, translated them into English on the basis of instructions provided, and provided preliminary ratings. Two international scoring sessions were arranged to rate



**Figure 1.** Design of the Writing Study  
(Gorman, Purves & Degenhart, 1988, p.10)

the scripts in teams. This provided feedback so countries could see how closely the international juries agreed with their ratings. This procedure also produced benchmarks that could be used as supports in national rating sessions.

### **Development of Conceptual Models for the Assessment of Writing: the Main Lasting Contribution of the Project?**

It seems to me that the main long-term contribution of the written composition study differs from a 'typical' IEA study. This study did not lead to sufficiently reliable and valid comparisons of the level of achievement among the participating countries, or to clear patterns that would explain the achievements. I believe that its main contribution was the development of quite novel and useful approaches to the assessment of writing (the first three research tasks), thus yielding a better conceptualization of the domain. The project was not only a huge challenge, but it was also an opportunity for raising awareness about the complexities of such an endeavor and about the strong influence of cultural and contextual factors in the teaching and assessment of writing.

While it is true that there was very useful literature to draw on (Cooper & Odell, 1977; Diederich, French & Carlton, 1961), the approaches used tended to be quite narrow in focus, building more on tradition than exploring new avenues. Our models for the domain of writing and for the assessment of writing were favourably received when the study was presented at several international conferences, including AERA (American Educational Research Association) and TESOL (Teachers of English to Speakers of Other Languages). I recall that one eminent reviewer referred to our work in print as “remarkable”. I will illustrate some of the models in the following section.

During the first months after the launch of the project, there was very intensive work devoted to developing models to get a better grasp of the construct of writing. A good example is the domain of writing developed by Vähäpassi and presented in Figure 2. It has become quite well known in in the mother tongue education profession.

Dominant intention/ Purpose	Cognitive Processing	Reproduce	Organize/Reorganize		Invent /Generate
	Primary Content Primary Audience	Facts Ideas	Events	Visual images facts,mental states,ideas	Ideas ,mental states alternative worlds
To learn (metalingual)	Self	copying, taking dictation	retell a story (heard or read)	note, resume, summary, outline, paraphrasing	comments on book margins metaphors analogies
To convey emotions, feelings (emotive)	Self, others	stream of consciousness	personal story personal diary, personal letter	portrayal	reflective writing personal essays
To inform (referential)	Others	quote fill in a form	narrative report news, instruction, telegram, announcement, circular	directions, description, technical description, biography, science report/experiment	expository writing definition academic essay/article book review commentary
To convince/persuade (conative)	Others	citation from authority/expert	letter of application statement of personal views opinions	advertisement, letter of advice	argumentative/ persuasive writing editorial critical essay/ article
To entertain, delight, please (poetic)	Others	quote poetry and prose	given an ending create a story create an ending retell a story	word portrait or sketch	entertainment writing parody, rhymes
To keep in touch (phatic)	Others	greeting card	postcard	personal letter	humorous greeting

**Figure 2.** Definition of the domain of writing in the Writing Study (Gorman, Purves & Degenhart, 1988, p. 22).



Figure 4 below). It also facilitated the development of a scoring scheme.

In the attempt to get a better grasp of the domain we also tried out Luis Guttman's approach to domain definition (e.g., Guttman, 1970), the facet analysis.

The approach proved interesting and indicated how vast the variety of written products could be. The model is not included in the published reports but appeared in the journal, *Research on the Teaching of English* (Purves, Söter, Takala & Vähäpassi, 1984).

In addition to the specification of the domain of writing, it was felt important to develop a model which would provide a conceptual basis for a rating scheme. There was no shortage of rating systems, beginning with the pioneering work by Diederich et al. (1961). Work done on "primary trait" scoring (e.g., Lloyd-Jones, 1977) was also a potential approach but it was not adopted as such. In primary trait scoring performances are evaluated by limiting attention to a single criterion (or a few selected criteria). This criterion or these criteria are based upon the trait or traits determined to be essential for the successful performance of a given task. For example, persuasiveness is the primary trait in an argumentative task. As far as I can recall, it was felt that it was too 'rigid' in demanding a closely genre-tied rating, whereas it was possible that not all cultures shared the same view of the appropriate realizations of different genres. Kaplan's early work on cross-cultural rhetoric (1966), as well as his subsequent work, also suggested the advisability of some caution in this respect.

Figure 4 presents the distribution of the nine different kinds of writing tasks across populations. The tasks will not be discussed in further detail. Suffice it to say that task 9, a letter of advice to a younger fellow student on how one should write in order to get good grades in the school, had a double function: it tapped the instructive text type but it also provided also an opportunity to analyze what views students had of what counts in writing at school. It was initially labeled "bizarre" by

Dominant intention/ Purpose	Primary Cognitive Demand		
	I Reproduce	II Organize Reorganize	III Invent Generate
1. To learn (metalingual, mathetic)		Summary (B,C) Paraphrasing (A)	
2. To convey emotions, feelings (emotive)		Narrative/Personal story (A,B)	Open essay (B,C)
3. To inform (referential)		Letter to an uncle describing a bicycle (A,B) Self description in letter to pen-pal (A,B) Formal note to head of school (A,B) Message to family (A) Application letter (B,C) Letter of advice to younger student (B,C) Describing an object (A,B) Describing a process (B,C)	Reflective essay (B,C)
4. To convince/ persuade (conative)	/	Application letter (B,C) Letter of advice to younger student (B,C)	Persuasive/ argumentative essay (A,B,C)
5. To entertain, delight, please (poetic)			Open essay (B,C)

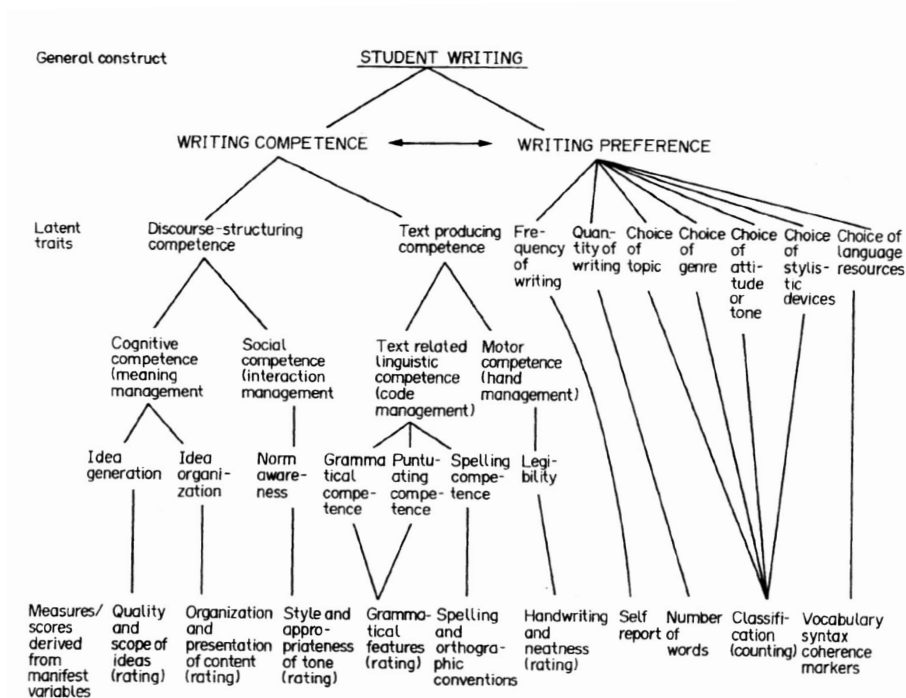
**Figure 4.** Distribution of tasks across Populations A, B and C  
(Gorman, Purves & Degenhart, 1988, p. 33)

one of the participants but turned out to be a very good idea. For a variety of reasons, it held a great appeal to students.

Instead, early work done by Canale and Swain (1980) on modelling communicative language ability, and by Lyle Bachman, who at that time was working on pioneering projects at UIUC (eg., Bacham &

Palmer, 1982), appeared promising. Combining these with the latent trait approach led to the development of the model for student writing represented in Figure 5. It breaks down the general construct of student writing into writing competence and writing preferences, and elaborates writing competence into the two latent traits of discourse-structuring and text-producing competence, and continuing such an elaboration to end up with manifest variables (rated scores).

The above description of model-building is not comprehensive. Similar work was done, for instance, on outlining different schemes of linking raters with each other to make it possible to estimate inter-rater agreement.



**Figure 5.** Model used to develop the general rating scheme (Gorman, Purves & Degenhart, 1988, p. 43).

## **Problems due to Inadequate Resources and Solutions Attempted**

The Spencer Foundation had initially provided funding to get the project off the ground (1981), but there was no success in raising further funding in the spring and summer of 1982. After a period of great anxiety, the Spencer Foundation extended its support by about two more years. Such support made it possible to hire a project coordinator (myself) at UIUC, an assistant/secretary, and later on part-time psychometric assistance. However, one of my most vivid recollections of the project is the constant worry over funding.

The project funding was running out again in the autumn of 1984 and I had to return to my home institute in Finland. I brought with me a massive dataset that needed to be analyzed to provide an international backdrop to national analyses. In the end, however, many countries had to write these reports without the international analyses. Understandably, this led to growing dissatisfaction in those countries that had a definite deadline for completing a national report.

The Institute for Educational Research at the University of Jyväskylä, Finland, was very helpful in providing data analysis services, initially under a brief contract with the IEA. However, this was a strain, as the data analysis resources were limited and there were numerous other projects to serve. Some time later, Elaine Degenhart arrived from Illinois and took on the main responsibility for practical coordination for the next three years. Although she was funded, this lasted for a very limited amount of time. Thus the situation was such that the coordinating center was now located in Finland but it had practically no resources allocated for the work. The participating countries, with a strictly pre-determined schedule and budgets, were clamouring for the international reports as a basis for their national reports. An emergency meeting was convened in Hamburg in early February in 1985 to try to find ways of making faster progress. Concrete plans were suggested and they were helpful, but as no major new funding was forth-



coming, faster progress was clearly not possible.

This called for creative fundraising: I applied for a vacant professorship in language didactics at the University of Helsinki and as I seemed to be the most qualified among the applicants I was asked to act as a pro tempore professor for a whole academic year. This released my salary and the Institute invested it in the IEA study. I subsequently also acted for a term as a substitute professor of applied linguistics at my own university and the released money was again used for the IEA study. We also managed to obtain a 3-month Fulbright scholarship for Alan Purves and Elaine Degenhart to come and work at the Institute in the summer of 1989. There was no summer holiday (a sacred thing in Finland and much needed after the hard winters!) during that summer; instead, there was intensive work even on weekends, focused on data analysis and writing up draft texts.

My home institute was by no means the only party to assist the project to push on despite the constant funding problems. For the project, vitally important was the support that members of the extended International Steering Committee managed to arrange in hosting meetings: Judit Kadarne-Fülop from the Hungarian National Institute for Education (OPI) hosted a meeting in Eszgeretom and Budapest, Thomas Gorman from the National Foundation for Educational Research in England and Wales (NFER) in Slough, Pietro Lucisano at the European Centre for Education (CEDE) in Frascati, Hildo Wesdorp from SCO in Amsterdam, Eva Baker at UCLA and Anneli Vähäpassi in Jyväskylä.

The chronic lack of funds created a number of problems:

- The international results were delayed and many national reports had to be written without having sufficient international comparative analyses to support their work (e.g., de Gopper, 1985; IEA IPS, 1988; Gubb, Gorman & Price, 1987; Lamb, 1987; Löfqvist, 1988).
- It was not possible to secure top statistical advice and services as the funding base was so insecure.

- The delays inevitably led to some tensions, occasionally quite serious, among the participants. Unreasonable amounts of work were required of many people under unreasonable deadlines.
- The use of partial least squares modelling (eg., Sellin, 1995) as one of main statistical methods for multivariate analyses was an extra challenge as it was less known in Finland than in Germany, requiring intensive correspondence.

The pressures we felt came also from inside the IEA itself. If my memory serves me, the reporting of the math study was much delayed and I recall that the vocal criticism for the delay caused pressures for the writing study to 'deliver' quickly, beyond what was at all reasonable given the problems in staffing due to the lack of funds. I remember objecting strongly to speeded-up reporting as this would have led to problems with quality.

The contributions of all the parties mentioned above made it possible to report the main outcomes of the study in two volumes (Gorman, Purves & Degenhart, 1988; Purves, 1992). As for the main findings, the following summary on the IEA website offers an apt and succinct account:

1. The construct "written composition" was found to be sited in a cultural context and so could not be considered a general cognitive capacity or activity. Marked variation across the countries existed in both ideology of the teachers and in instructional practices. Written performance was also found to be task dependent.
2. Good compositions from different countries shared common qualities related to handling of content and appropriateness of style, but these qualities had their national or local characteristics in organization, use of detail, and other aspects of rhetoric.
3. Students across educational systems had in common a sense of the importance of the written product and its surface features. Beneath that commonality, however, there was national variation in the perception of what is valued.
4. In most of the countries, girls were treated differently than boys in the

provision of writing instruction and in the rating of writing performance, particularly at the primary and lower secondary school levels where women largely provided instruction. In such a milieu, the most successful students were girls, and gender itself, or gender in combination with certain home variables, was the most powerful predictor of successful performance, particularly on the more 'academic' tasks.

5. Differences between the ratings of student writing were not explained by differences in instruction. They were, however, accounted for by factors involving the characteristics of the home, the reinforcement provided by parents, and the cultural values of the community.

[http://www.iea.nl/written\\_composition.html](http://www.iea.nl/written_composition.html) (accessed July 18 2011)

### **Impact of the Study**

I have already indicated that, for most of the participating countries, the written composition study represented a new venture. As in the long history of the IEA's work, participation in its projects offered hands-on training in how large-scale assessments of writing could be carried out. I will give a couple of specific examples I am familiar with.

In Italy, the very extensive curriculum and teacher questionnaires were used as a basis for organizing in-service training for teachers. These tools were felt to be very useful for raising awareness of the range of options in teaching and assessing writing. In Finland, the Finnish national coordinator, Anneli Vähäpassi, invited all teacher education departments in Finland to send representatives to the national ratings sessions; several of these participated in writing chapters in Finnish national reports. She also developed booklets drawing on the IEA tasks and their rating schemes, and she made these available to schools/teachers on a true-cost basis.

The study brought together a large number of mother tongue specialists from all over the world and helped to establish much-needed international networks. For me personally, the opportunity to work closely with such an eminent scholar as Alan Purves was a source of

constant inspiration leading to the most productive phase in my career. In addition to our joint efforts for the project, we often discussed the nature of international comparative research. On the request of the IEA Chairman, Neville Postlethwaite, we conducted a survey of what the participating countries saw as the main advantages of conducting cross-national surveys of educational achievement. Alan referred to this in an article (1987) in the *Comparative Education Review* (included in this book).

The initiative to establish an international essay database that would allow secondary analyses might have had a potentially great impact had it been successful. Some countries sent their scripts to the putative centre, but again, funding did not materialize and no such 'clearing-house' was established. One national coordinator noted that the scripts sent in simply vanished, never to be seen again.

The study gave rise to a number of conference presentations and publications over and above the international reports. Articles were published in a number of prominent journals, including the *Comparative Education Review* (Takala & Vähäpassi, 1987), *Evaluation in Education* (Purves & Takala, 1982), *Written Communication Annual* (edited by Purves, 1988).

As noted above, the study was accompanied by considerable pressure and anxiety over a number of years, and a frightening personal experience of lack of sleep over several days due to a very bad jet lag. However, these were more than compensated by becoming acquainted and working with great colleagues all over the world. Life with the study was sometimes quite exhausting, occasionally quite frustrating, but always exciting, interesting and instructive.

## References

- Bachman, L. F., & Palmer, A. S. (1982). The construct validation of some components of communicative proficiency. *TESOL Quarterly* 16 (4), 449-465.

- Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics* 1 (1), 1-47.
- Carroll, J. B. (1975). *The teaching of French as a foreign language in eight countries*. Stockholm: Almqvist & Wiksell.
- Cooper, R. C., & Odell, L. (Eds.). (1977). *Evaluating writing: describing, measuring, judging*. Urbana, IL: National Council for Teachers of English.
- de Glopper, K. (1985). *Schrijprestaties in het voorgezet onderwijs. Een onderzoek naar de stelvaardigheid van leerlingen in het der der de leerjaar LTO, LNHO, MAVO, HAVO en VWO* [Writing performance in secondary schools]. Stichting Centrum voor Onderwijsonderzoek van het Universiteit van Amsterdam. Den Haag: SVO.
- de Glopper, K. (1995). Instruction and assessment in written composition in primary and secondary schools. *European Education* 27 (1), 71-86.
- Degenhart, E. & Takala, S. (1988). Developing a rating method for stylistic preference. A cross-cultural pilot study. In A.C. Purves (Ed.), *Writing across cultures. Issues in contrastive rhetoric*, (pp. 79-106). Newbury Park: Sage.
- Dickson, P. & Cumming, A. (Eds.). (1996). *Profiles of Language Education in 25 Countries*. UK: National Foundation for Educational Research.
- Diederich, P. B., French, J. W., & Carlton, S. T. (1961). *Factors in judgments of writing ability. Research Bulletin 15*. Princeton, NJ: Educational Testing Service.
- Gorman, T. P., Purves, A. C., & Degenhart, R. E. (1988). *The IEA Study of Written Composition I: The International Writing Tasks and Scoring Scales*. Oxford: Pergamon Press.
- Gubb, J., Gorman, T., & Price, E. (1987). *The study of written composition in England and Wales*. Windsor: NFER-Nelson.
- Guttman, L. (1970). Integration of test design and analysis. In "Toward a theory of achievement measurement". *Proceedings of the 1969*

- Invitational Conference on testing problems*, pp. 53-61. Princeton. NJ: Educational Testing Service.
- Hartmann, W. (1995). Different Approaches to Comparative Mother-Tongue Education (CME). In W. Bos & R. Lehmann (Eds.), *Reflections on educational achievement*, (pp. 119-136). Münster: Waxman.
- Herrlitz, W., Ongstad, S., & van den Ven, P. (Eds.). (2007). *Research on mother tongue education in a comparative perspective. Theoretical and methodological issues*. Amsterdam: Editions Rodopi.
- IEA IPS. Indagine sulla produzione scritta: Italia. *Ricerca educativa*, 1988, (2-3), 1-251.
- Kaplan, R. (1966). Cultural thought patterns in intercultural education. *Language Learning* 6 (1), 1-20.
- Lamb, H. L. (1987). *Writing performance in New Zealand schools. A report of the IEA study of written composition in New Zealand*. Wellington: Department of Education.
- Löfqvist, G. (1988). *Elever skriver uppsats*. [Pupils write composition.] Lund: Studentlitteratur.
- Lewis, E. G., & Massad, C. E. (1975). *The teaching of English as a foreign language in ten countries*. Stockholm: Almqvist & Wiksell.
- Lindell, E. (1980). *Six reports on free writing: A summary of the Fris project*. Malmö School of Education. Lund Universitet: Department of Educational and Psychological Research, Report No. 61.
- Lloyd-Jones, X. (1977). Primary trait scoring. In C. Cooper, & L. Odell (Eds.), *Evaluating writing: Describing, Measuring, judging*, (pp. 33-66). Champaign. Ill: National Council for Teachers of English.
- Odell, L. (1982). Defining and assessing competence in writing. In C. R. Cooper (Ed.), *The nature and measurement of competency in English*, (pp. 95-138). Urbana, Il.: National Council of Teachers of English.
- Petrovsky, A. R., & Bartholomae, D. (Eds.). (1986). *The teaching of writing. Eighty-fifth yearbook of the National Society for the Study of Education, Part II*. Chicago: University of Chicago Press.

- Purves, A. C. (1973). *Literature education in ten countries*. Stockholm: Almqvist & Wicksell.
- Purves, A. C. (1979). *Achievement in reading and literature in the secondary schools. New Zealand in an international perspective*. Wellington: New Zealand Council for Educational Research.
- Purves, A. C. (Ed.). (1979). *Papers presented at the Pittsburg symposium on cognition and written language*. March 15. ERIC ED 178918.
- Purves, A. C. & Takala, S. (Eds.). (1982). An international perspective on the evaluation of written composition. *Evaluation in Education: An International Review Series* 5 (3).
- Purves, A. C., Söter, A., Takala, S. & Vähäpassi, A. (1984). Towards a domain-referenced system for classifying composition assignments. *Research in the teaching of English* 18, 385-416.
- Purves, A. C. (1987). The evolution of the IEA: A memoir. *Comparative Education Review* 31 (1), 10-28.
- Purves, A. C. (Ed.). (1988). *Writing across cultures. Issues in contrastive rhetoric*. Newbury Park: Sage.
- Purves, A. C. (1992). *The IEA Study of Written Composition II: Education and Performance in Fourteen Countries*. Oxford: Pergamon Press.
- Purves, A. C. (1992). Reflections on research and assessment in written composition. *Research in the Teaching of English* 26, 108-122.
- Sellin, N. (1995). Partial least squares modeling in research on educational achievement. In W. Bos & R. Lehmann (Eds.), *Reflections on educational achievement*, (pp. 156-267). Münster: Waxman.
- Smith, L. S. (1978). *Investigation of writing assessment strategies*. UCLA: Center for the Study of Evaluation.
- Squire, J. R., & Applebee, R.J. (1969). *Teaching English in the United Kingdom. A comparative study*. Champaign, Ill; National Council of Teachers of English.
- Takala, S. (1983). *Student views of what counts as a good composition: some national findings*. Paper presented at the 7th Annual TESOL Conference, Toronto.

- Takala, S., & Vähäpassi, A. (1987). Written communication as an object of comparative research. *Comparative Education Review* 31, 88-105.
- Thorndike, R. L. (1973). *Reading comprehension education in fifteen countries*. Stockholm: Almqvist & Wicksell.
- Weigle, S. C. (2002). *Assessing writing*. Cambridge: Cambridge University Press.
- Wesdorp, H. (1978). *Evaluatietechnieken voor het moedertaalonderwijs. Een inventarisatie van beoordelingsmethoden voor de stelvaardigheid, het begrijpend lezen, de spreek-, luister- en discussie-vaardigheid* [Evaluation techniques in mother tongue instruction]. Amsterdam: Uva/RTP





## CHAPTER 6

# **IEA Assessments of Information and Communication Technologies (ICT) in Education**

*Willem J. Pelgrum & Tjeerd Plomp*

### **Introduction**

Throughout the past three decades, information and communication technologies (ICT) have had a major impact on all societies. Unprecedented changes have taken place, not only in the labour market in terms of job qualifications and trading processes, but also in the day-to-day activities of most citizens. Many traditional ways of communication, industrial production, banking and the like, have been replaced by digital technologies.

It was some time in 1985 that the IEA acknowledged that the informatization of societies would have major consequences for education as well. At the same time, many educational actors (policy makers, educationalists, educational practitioners) were speculating about the role of ICT in education. Yet there was very little data on the use of ICT in daily educational processes and the extent to which national developments compared to developments abroad. As a result, and in order to empirically document these developments, IEA began to make international comparative assessments. To date, the following assessments have been conducted and results have been disseminated:

- The Computers in Education Study-stage 1 (COMPED-I) - data were collected in 1989 from national samples of schools and teachers in 21 education systems.
- The Computers in Education Study-stage 2 (COMPED-II), in which, in 1992, national samples of schools, teachers and students from 12 education systems participated.
- The Second Information in Education Study- module 1 (SITES-

M1), with data collection in 1999 among representative samples of school leaders and technology coordinators from 26 education systems.

- The Second Information in Education Study- module 2 (SITES-M2), consisting of case studies of innovative practices utilizing ICT that took place in 28 education systems in 2002.
- The Second Information in Education Study 2006 (SITES-2006), which collected data from schools and teachers in 22 education systems in 2006.

Some characteristics of each of these studies as well as the key actors, are presented in Appendix 1.

Currently (2011) the International Study of Computer and Information Literacy (ICILS), for which data will be collected in 2013, is underway.

Most of the above-mentioned studies are quantitative international assessments. An exception was SITES-M2, which was qualitative in nature and consisted of 174 standardized case studies in schools that were forerunners in applying ICT in their education system or country.

The aim of this chapter is to offer a global overview of the completed studies.

This will be done by describing first the main research questions and conceptual frameworks that were used in these studies and by characterizing how these frameworks changed over the years as a result of experiences of the early adopters and the developing perceptions of educational stakeholders about the potential role of ICT. This will be followed by a bird's eye-overview of major observations resulting from the analysis of the data from these studies. Finally, reflections regarding the future of ICT-related assessments will be offered.

## **Research Questions and Conceptual Frameworks**

As can be inferred from the above, the IEA ICT assessments have been

carried out for over 20 years. Over these years, quite naturally, the educational context and the expectations regarding the role of ICT in education have changed. This had an impact on research questions that were addressed in these studies, conceptual frameworks that were used to address these questions and issues that played a role in the design of the assessments. For instance, when COMPED-I was developed, there were still substantive numbers of schools in industrialized countries that did not have access to computers, or in those schools that did have computers, teachers and/or students did not have access. Therefore an important issue at that time was to investigate whether schools had computers and how many, as well as why some schools (or teachers and/or students) used computers while others did not. The design of the study had to reflect the distinction between users and non-users. Currently this issue hardly plays a role anymore at school level in economically developed countries: SITES-2006 showed that all schools in the participating countries except South Africa (see Appendix 1 for an overview of countries) possessed an ICT infrastructure for teaching and/or learning purposes.

Another major change has been that the characteristics of the equipment (and consequently the kind of information that needed to be collected about the available equipment) have changed substantially, such as from micro-computers to multimedia computers, from stand-alone to networked, from floppy disks to hard disks.

However, a rather constant conceptual perspective that has been prevalent in all studies was the focus on theories of educational change. It was acknowledged that the introduction of ICT in education should be considered a major educational change, and that the success of implementation and the impact on educational processes and outcomes depend on a large number of non-ICT factors such as the pedagogical vision of educational practitioners, school leadership, and continuous staff development facilities. Moreover, it was acknowledged that quite a number of school-external factors played a role in the way ICT had been adopted and implemented for educational purposes.

The measures used in the COMPED assessments were based on a conceptual framework characterizing the educational system in terms of decision-making at different levels: the macro- (national), meso- (school) and micro- (class). The framework identified a number of factors that were assumed to contribute to effecting changes, which are shown below in italics:

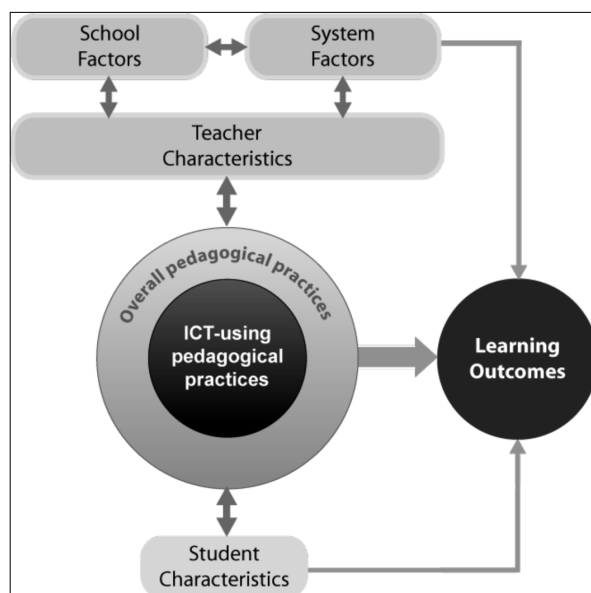
- It is necessary that the objectives of an innovation are perceived as *clear* and *relevant* by educational practitioners who are involved in its implementation.
- Materials used in the innovation process (manuals, guidelines, teaching materials) should be of high *quality*.
- Continuous *support* for solving day-to-day problems as well as strong *leadership* within schools are also important conditions for stimulating teachers to adopt and implement the intended changes.
- Adequate facilities for *training* teachers and *continuous staff development* are necessary for teachers to learn how to translate the intended objectives into daily lesson practices.
- Finally, a system of *continuous evaluation* and the provision of *feedback* to actors involved in the innovation process at different levels in the educational system are important for monitoring the pace and direction of the changes.

The framework reflected the hierarchical structure of most educational systems, but also acknowledged that decisions which promote or inhibit the implementation of computer-related curricula and applications are made at all levels, which may cause discrepancies between decisions and expectations that exist at different system levels. An identification of these discrepancies was seen as a potential important starting point for initiating improvement measures in education.

When micro-computers were first introduced in education, non-ICT factors were given scant attention. Hence, the IEA studies, for instance,

repeatedly underscored the tension between existing curriculum guidelines and the need for greater flexibility, which educational practitioners experienced as a necessary condition for integrating ICT in teaching and learning. The awareness among educational practitioners that the integration of ICT in education is a complex educational innovation has clearly grown over years. It seems safe to assume that IEA publications (and the many presentations of these at conferences throughout the world) have contributed to this increased awareness, although we can only speculate on the extent of this influence. But it is interesting to observe that while in the mid-1980s it was common to hear it said that, “ICT will lead to more productive learning among students,” now there has been a shift in emphasis to statements like “ICT can play a facilitating role in realizing well-defined educational visions”.

An overall conceptual model reflecting the above-mentioned factors and which underlay all IEA ICT assessments was presented in the final report of SITES-2006 (Law, Pelgrum & Plomp, 2008; see Figure 1).



**Figure 1.** Overall conceptual framework for the SITES 2006 study (Law, Pelgrum & Plomp, 2008)

As an illustration, Table 1 shows how a concept like 'school-level factors' was elaborated in SITES-2006 and ultimately operationalized in questionnaire items for school leaders.

**Table 1.** Summary of the contents of the school questionnaires (Law, Pelgrum & Plomp, 2008)

Concepts addressed in school questionnaires	Description
Infrastructure	Availability of ICT hardware (types of computers, local area network, Internet connections, electronic whiteboards, etc.) and software (general and subject specific software, learning management systems, assessment tools, etc.), infrastructure needs, problems
Pedagogical practice	Extent to which Life Long Learning practices are present in the school
Vision	The vision of the school management with regard to pedagogy and ICT covering three dimensions: traditional, Life Long Learning (LLL) and Connectedness.
Staff development	Encouragement or requirements for teachers to acquire knowledge and skills with regard to pedagogical practices and the use of ICT. Priorities for acquiring competencies by the school leadership Ways that teachers in the school have acquired knowledge and skills for using ICT in teaching and learning Availability (school-based and/or externally) of ICT-related courses
Support	Persons involved in providing support and time-expenditure Extent to which pedagogical support is available for teachers Extent to which technical support is available for teachers
Organization and management	Role of principals in initiating changes Decision-making responsibilities Management of change Stimulation of cooperation between teachers Promotion of alternative assessment practices

## Major Observations from the Different Studies

Since the early days of using computers in education, educational stakeholders have been asking questions about the (potential) impact of ICT on educational processes and outcomes, as well as about factors influencing or impeding the pedagogical use of ICT. The aim of the IEA ICT studies has always been to find empirical answers to these questions. Some of these questions are addressed below and a summary of insights afforded by the different studies is presented.

The main questions that have been asked throughout the years about technology and education deal with access, use, knowledge and attitudes of students, support, pedagogy, staff development and obstacles. Some findings on each of these will be presented.

### *Access*

During the early days of ICT use in education, the first questions dealt with access: do schools, teachers and students have, in principle, access to the new technology? Is the available infrastructure adequate and up-to-date? Although such questions can rarely be answered in an absolute sense, the IEA ICT assessments offered educational stakeholders a relative answer, showing how countries and even groups within countries differed in terms of access to technology. In COMPED-I (data collected in 1989) it was shown (Pelgrum & Plomp, 1991) that in some countries substantial groups of schools had no access to computers, while in other countries all schools already were equipped. While today there are only a handful of schools in the developed world lacking computer equipment for instructional purposes, the quantity of available hardware differs - sometimes dramatically - among as well as within countries. Such inequities are important signals for policy makers who strive to offer equitable conditions throughout their educational systems.

At the time of COMPED-I most educational practitioners (computer coordinators, principals and teachers) perceived a shortage of hard-



ware and software as the two main problems associated with introducing computers in the school curriculum; computer coordinators saw the acquisition of a greater variety of software as the highest priority. It was concluded in 1991 by the researchers that the critical mass of computers and software needed for a proper integration of computers into the curriculum was not yet reached; in other words, in 1989 there was, generally speaking, still an insufficient basic infrastructure in schools for using computers. COMPED-II revealed in 1993 (Pelgrum & Plomp, 1993) that the availability of hardware and software in schools was slowly increasing, but that, except for the USA, substantial groups of students still did not seem to have access to computers at school.

SITES-M1 also examined (in 1998) the extent to which schools had access to the Internet for instructional purposes. Again, there were significant differences between countries (Pelgrum & Anderson, 1999). In Singapore and Iceland, 100% of the lower secondary schools had access, while this figure was 98% in Canada and 96% in Finland. On the other hand, 11% of schools in Cyprus and only 4% of the Russian lower secondary schools had at that time access to the Internet.

Between SITES-M1 (1999) and SITES-2006, school access to the Internet improved substantially in most countries (Law, Pelgrum & Plomp, 2008).

### *Use of ICT*

Although access to technology is an essential pre-condition for instructional use, the IEA assessments repeatedly showed that computer use for instructional purposes was for many years a rather marginal activity during lessons. Students mainly used computers outside school. Even in SITES-2006 it was shown that during the majority of lessons in science and mathematics ICT was hardly used, although it was also observed that between 1998 and 2006 pedagogical practices reflected a considerable increase in information handling activities.

Interesting to note in COMPED I (in 1989) was the existence of phantom users: students using computers in schools that did not possess computers. This could happen when schools, for instance, let their students visit locations other than the school where they could use computers or when students wrote computer programs at school that were sent to a computer centre for processing and then returned to the school with a print of the program output.

Based on the data from 1989, it was concluded that in secondary education, computers were used primarily as an add-on to the already existing curriculum in the form of teaching students how to use the computer. The most common practice in lower and upper secondary education was to offer this kind of instruction as a separate course. Most times that such a course did not exist, computer education was taught as part of mathematics. In 1989 integration of the computer into the existing subjects of secondary schools was still in its initial stage. In elementary education, computers were most frequently used for drill and practice, while in secondary schools word processing and programming were most popular. This situation did not change much between 1989 and 1992. Although at the school level a gradual increase of computer use in existing subjects could be observed in some countries, students still used computers very infrequently in the context of these subjects.

The data collected in SITES-2006 led to the conclusion that increasing the level of computer access per se does not result in more learning experiences that improve students' learning outcomes.

### *Knowledge and Attitudes of Students*

Students' knowledge was measured in COMPED-II (in 1992) by a 'functional information technology' test (FITT). The FIT-Test was designed from the perspective of 'functionality', that is, to measure essential computer-related abilities which allow students to function effectively on information-related tasks. The results suggested that fac-

tors other than formal teaching in schools contributed substantially to knowledge about new information technologies, such as using computers outside the school or simply learning by doing. Student knowledge and student attitudes were quite different and seemed to relate to the amount of computer exposure. For instance, the highest scores were for students who used computers both at school as well as outside, but those who used computers only outside school had higher scores than those who used computers only at school. Most students perceived computers as relevant for their future, but did not always enjoy computer-related activities. In a number of countries the encouragement by parents to use computers was relatively low. It was observed (in 1992) that quite a number of students tended to agree with unethical practices like the illegal copying of software, which raised the question to what extent schools and parents should play a more prominent role in making students more aware of the implications of such practices.

SITES-M2 showed that certain patterns (of computer practice) were more likely to be associated with significant positive outcomes (as perceived by those who were interviewed during the case studies). For example, in the cases where technology supported students to collaborate with one another, to conduct research, and to analyze data, the respondents were far more positive about students having acquired new ICT, problem-solving, and collaboration skills than respondents with other practice patterns in using ICT.

SITES-2006 demonstrated that the impact of ICT use on students appeared to be highly dependent on the pedagogical orientation that teachers adopted in regard to that use. Data analysis revealed significant positive correlations between lifelong-learning-oriented pedagogical uses of ICT in teaching and learning and perceived gains in students' 21st-century outcomes (for an illustration of these concepts, see below under 'Pedagogy'). No significant correlations were found between traditionally oriented uses of ICT and students' learning outcomes, as reported by their teachers.

### *Support*

From the case studies of innovative pedagogical use of ICT in SITES-M2 it was inferred (Kozma, 2003) that innovations were more likely to continue after initial use, if there were support from others in the school and also from external sources, innovation champions, funding, and supportive policies and plans. Particularly important was a connection with national technology plans that provided resources that often enabled the innovation to succeed.

SITES-2006 offered further evidence that the most serious obstacles to ICT utilization in the classroom are school related rather than student related. The SITES data showed that the surveyed teachers identified lack of support as the most significant obstacle.

The most important school-level factors influencing teachers' ICT use for lifelong-learning practices are the vision that principals had in regard to ICT use supportive of lifelong-learning pedagogy and the technical and pedagogical support available to teachers and students. These findings held for systems with different mean lengths of ICT-using experience in their schools. This may imply that late starters are not necessarily lagging behind the early adopters.

### *Pedagogy*

Perhaps the most significant aims of SITES-M1 were to examine the extent to which countries were changing their approach to pedagogy and to look at the contribution that ICT was making to this change. Principals were asked a number of questions about the presence of pedagogical practices in their schools. The list of questions appears in Box 1. A factor analysis was run on the responses to these questions and two factors were created: one called emerging practices and the other called *traditional practices*. The *emerging practice factor* was formed from items 1, 2, 3, 4, 8, 9, 10, and 13 in Box 1, while the *traditional practice factor* was formed from items 5, 6, and 7. In brief, the emerging practices included those that made students active in and responsible

for their own learning, that involved students in cooperative or project-based learning, that engaged students in the search for information, and that allowed them to work at their own pace and determine when to take a test. Traditional practices were those that emphasized the development of skills, where all students were working on the same materials at the same pace, and where teachers kept track of all student activities and progress.

**Box 1:** Questionnaire item about pedagogical practices

**Question:** To what extent is each of the following aspects of teaching and learning present in your school?

*(Response alternatives were: not at all, to some extent, a lot, for each of the following practices.)*

1. Students developing abilities to undertake independent learning
2. Providing weaker students with additional instruction
3. Organizing teaching and learning so that differences in entrance level, learning pace, and learning route are taken into account
4. Students learning to search for information, process data, and present information
5. The emphasis in learning is on the development of skills
6. Students working on the same learning materials at the same pace and/or sequence
7. Teachers keeping track of all student activities and progress
8. Students being largely responsible for controlling their own learning progress
9. Students learning and/or working during lessons at their own pace
10. Students involved in cooperative and/or project-based learning
11. Students determining for themselves when to take a test
12. Students learning by doing
13. Combining parts of school subjects with one another (multidisciplinary approach)

Many schools around the world indicated that the emerging pedagogical practices were present to a large extent in their schools. However, as with other indicators, there were also large differences between countries in their pedagogical practices. For example, on a scale with a score range of 0-100, Norway scored the highest with 71, and Denmark and Hungary scored 69 on emerging pedagogical practices in their lower secondary schools. China- Hong Kong scored 36 and Japan scored 29, the lowest. Regarding traditional practices, Thailand scored 75 and Luxembourg scored 72, while Norway scored 43. Beyond this, principals were asked the extent to which ICT had contributed to the realization of the various emerging pedagogical practices in their schools. In Denmark, 62% of the lower secondary principals responded affirmatively; the figure was 58% in Israel and 56% in Canada, Hungary, and Slovenia. The figure was much lower in China-Hong Kong (40%), the French Belgium community (37%), and Japan (31%).

From SITES-M2 it was inferred that in a substantial number of cases, technology was supporting significant changes in classroom teaching and learning. These cases painted a very different picture from that of the traditional classroom where the teacher lectures in front of the classroom and students take notes or do worksheets. The changes also showed important similarities in the manner in which innovative schools in many countries around the world were using technology.

It was also inferred that successful innovation did not spread easily: while 75% of the innovations had been used for at least a year, only 41% provided evidence that the innovation had been disseminated to other classrooms or schools. In the schools where these innovations had been both continued and disseminated, continuation depended on the energy and commitment of teachers, student support, the perceived value of the innovation, the availability of teacher professional development opportunities, and administrator support.

The emphasis on autonomous learning of students increased between

SITES-M1 and SITES-2006 considerably in some systems (e.g., Hong Kong), but there was also a noteworthy decrease observed in other education systems (e.g., Denmark), that had scored quite high on this indicator during SITES-M1 but in the SITES-2006 study showed a change towards a more traditional paradigm (with a focus on curriculum-centred teaching and instruction).

SITES-2006 also showed that teachers' pedagogical orientations, such as the degree of their understanding of the changing demands of citizens in the knowledge economy and their readiness to employ more collaborative, inquiry-oriented learning activities, to create a more open and connected learning environment, and to take on more facilitative roles, made a major difference to the way teachers utilized ICT in their classrooms.

### *Staff Development*

In 1989 it was observed that the amount of teacher training in existing subjects and the type of topics covered were related to the degree of computer integration. An important finding was that while teachers indicated that being trained specifically in pedagogical/instructional aspects of computer use was relevant, in fact, these topics were those least covered in teacher training programs until 1989. Throughout the IEA ICT assessments it appeared that continuous staff development was a problematic issue in many countries (see below under 'Obstacles').

### *Obstacles*

From SITES-M1 (1998) it appeared that, despite the general increase in the availability of computers and their connection to the Internet, the problem most often mentioned by respondents was the insufficient number of computers (as was found in the COMPED studies). Other infrastructure-related problems often mentioned by respondents included insufficient peripherals, not enough copies of software, and insufficient number of computers that could simultaneously access the Internet.

However, the second most-mentioned problem was teachers' insufficient knowledge and skills in ICT. While the majority of schools reported having a policy goal of training all teachers in ICT use, only a minority of schools in most countries were successful. As for the computer coordinators, or those persons who answered the technical questionnaire, a majority across countries responded that they were adequately prepared with regard to general applications (such as word processing, database and spreadsheet software). But the percentage was much lower with regard to the instructional aspects of ICT (for instance, didactical integration and applications of subject specific software).

### **The Current IEA ICT-Assessment: ICILS**

At the time of writing this chapter, the next IEA study on ICT in education had just begun. This study is called the International Computer and Information Literacy Study (ICILS), and it will examine the outcomes of student computer and information literacy (CIL).

The following definition of Computer and Information Literacy (CIL) is the basis of the proposed study.

*Computer and information literacy refers to an individual's ability to use computers to investigate, create and communicate in order to participate effectively at home, at school, in the workplace and in the community.*

The main research questions of this study are:

1. What variations exist between countries, and within countries, in student CIL?
2. What aspects of schools and education systems are related to student achievement in CIL?
3. What characteristics of students' backgrounds, levels of access to, familiarity with and self-reported proficiency in using computers are related to student achievement in CIL?



4. What aspects of student personal and social background (such as gender, socio-economic background, and language background) and familiarity with computers are related to CIL?

The assessment of CIL at student level will be authentic and computer based. It will incorporate three types of items (or tasks): 1) multiple-choice or constructed response items based on realistic stimulus material; 2) software simulations of generic applications so that students are required to complete an action in response to an instruction; and 3) authentic tasks that require students to modify and create information products using 'live' computer software applications. In addition, a number of associated instruments will be administered on the use of computers by the students, teachers, school principals and information technology coordinators.

The project would aim to develop the pilot materials during the year 2011, conduct a field trial in the first half of 2012, collect data at the beginning of 2013 (northern hemisphere), with reporting in May 2014. More information about this new study can be found at: [www.iea.nl](http://www.iea.nl).

### **Reflections on Future International Comparative ICT Assessments**

A general conclusion from COMPED was that the application and integration of computers in education is a very complicated process, expensive and beset with problems, requiring much time investment by educational practitioners. Moreover, it was argued that setting goals in this field is very complicated due to the fact that hardware and software applications are in a constant state of flux.

The findings of the IEA studies on ICT in education have been used (in combination with other international assessments on this domain) in various ways. Examples include an EU study on ICT indicators in education ([http://eacea.ec.europa.eu/llp/studies/study\\_on\\_indicators\\_on\\_ict\\_education\\_en.php](http://eacea.ec.europa.eu/llp/studies/study_on_indicators_on_ict_education_en.php)), and a US Department of Education study entitled

“International Experiences with Technology in Education” (to be published by the Education Publications Center, US Department of Education).

As the first author of this chapter was coordinating the EU project, a number of reflections from this study are summarized here point by point, illustrating the relevance of IEA type studies on ICT in education.

- Policy makers need educational monitors in order to make inferences about the strengths and weaknesses in the competencies of students, how these are developing over time and what are the potential causes of weaknesses. This holds for education in general, but also more specifically for ICT. With regard to ICT the core questions which should be addressed by educational monitoring are: (1) during the years of compulsory education, are students sufficiently skilled to use ICT in the competency areas targeted by the European Commission for benchmarking; and (2) do students have sufficient opportunities to learn about ICT at and/or outside school? Indicators are needed to address these questions.
- A survey among ICT policy experts in the EU showed a high need for indicators regarding ‘Opportunities to learn with and/or about ICT’, ‘ICT-related competencies and attitudes of students’, ‘ICT support’, ‘Teacher training’, and ‘School leadership’
- Suitable indicator definitions do not exist for the key competency areas that should (ideally) constitute the core for monitoring ICT in education, namely the ICT-related student outcomes and opportunities to learn.
- International comparative indicators of student competencies in the EU, which are regularly collected, exist for only a few traditional subject areas. ICT-related indicators collected from students mainly concern very general indicators of computer use in and/or outside school. Moreover, for those indicator areas that are covered in existing international comparative assessments, there are

data gaps: for many countries the time series studies since 2000 are incomplete or non-existent. Hence, new indicators definitions and instruments need to be developed.

- Ideally, monitoring should lead to a well-founded evaluation of the strengths and weaknesses of ICT use in educational systems. It was concluded that, in order to avoid undesirable side effects, a broad coverage of competency areas is needed. Next to traditionally valued competences, so-called 21st century skills should also be covered (see above under 'Pedagogy').
- Previous assessments contain many examples of indicator definitions for explanatory indicators. It was concluded that for future use these definitions need to be fine-tuned to definitions of the main indicators (regarding student competencies and the opportunity to learn about ICT).

The conclusions arising from the EU study illustrate that there is a need for monitoring the use and impact of ICT internationally. We believe that for developing a regular monitor on ICT in education, an organizational model like the one utilized in the IEA studies has proven to be effective. The organization of these studies comprises an international coordination centre, national coordination centres, international committees of experts for developing assessment frameworks and instruments for measuring student competencies in the core areas, and national expert committees in which educational actors from different organizations are represented (ministry, school inspection, school leaders, teachers, parents).

## References

### COMPED-1

Pelgrum, W.J., & Plomp, T. (1991). *The use of computers in education worldwide: results from the IEA 'Computers in Education' survey in 19 education systems*. Oxford: Pergamon Press.

**COMPED-2**

Pelgrum, W.J., & Plomp, T.(Eds.). (1993). *The IEA study of Computers in Education: implementation of an innovation in 21 education systems*. Oxford: Pergamon Press.

Pelgrum, W.J., Janssen Reinen, I., & Plomp, T. (Eds.). (1993). *Schools, Teachers, Students and Computers: a Cross-National Perspective*. The Hague: International Association for the Evaluation of Educational Achievement.

**SITES-M1**

Pelgrum, W.J., & Anderson R.E. (Eds.) .(1999). *ICT and the Emerging Paradigm for Life-Long Learning*. Amsterdam: IEA.

**SITES-M2**

Kozma, R.B. (Ed.) (2003) *Technology, innovation and educational change: a global perspective*. Eugene: ISTE.

**SITES-2006**

Law, N., Pelgrum, W.J., & Plomp, T. (Eds.). (2008). *Pedagogical practices and ICT use around the world: findings from an international comparative study*. CERC Studies in Comparative Education. Hong Kong: Comparative Education Research Centre, The University of Hong Kong, and Dordrecht: Springer.

**ICILS [www.iea.nl](http://www.iea.nl) and <http://forms.acer.edu.au/icils/index.html>**

## APPENDIX 1. Study characteristics

### IEA-study: Computers in Education, Stage I

**Abbreviation:** COMPED-I

**Start:** 1987

**End:** 1990

**Year of Data Collection:** 1990

**Main study report:** Pelgrum, W.J. & Plomp, T. (Eds.). (1993). *The IEA study of Computers in Education: implementation of an innovation in 21 education systems*. Oxford: Pergamon Press.

#### Staff

*International Coordination Center*

Willem J. Pelgrum (International Coordinator)

*International Steering Committee*

Tjeerd Plomp (Chair), Richard Wolf, Ryo Watanabe

*Sampling Referee*

Colm O'Muircheartaigh

*Data Management*

Leendert van Staalduinen, Rien Steen

#### Education systems:

Austria	Italy
Belgium-Flemish	Japan
Belgium-French	Luxembourg
Canada-British Columbia	Netherlands
China	New Zealand
France	Poland
Germany	Portugal
Greece	Slovenia
Hungary	Switzerland
India	United States of America
Israel	

#### Populations (elementary, lower secondary, upper secondary)

- Non-using schools
- Using schools
- Computer-using teachers (for secondary education: broken down by existing subjects -math, science and mother tongue - and computer

education)

- Non-computer-using teachers (for secondary education: broken down by existing subjects -math, science and mother tongue)

**Instruments**

- School Questionnaire (principal part)
- School Questionnaire (technical part)
- Teacher Questionnaire Computer Education
- Teacher Questionnaire Existing Subjects (international option)
- Student background characteristics (age, gender, SES, etc.)

## IEA-study: Computers in Education, Stage II

**Abbreviation:** COMPED-II

**Start:** 1990

**End:** 1993

**Year of Data Collection:** 1992

**Main study report:** Pelgrum, W.J., Janssen Reinen, I., Plomp, T. (Eds.). (1993). *Schools, Teachers, Students and Computers: a Cross-National Perspective*. The Hague: International Association for the Evaluation of Educational Achievement.

### Staff

#### *International Coordination Center*

Willem J. Pelgrum (International Coordinator), Ingeborg Janssen Reinen

#### *International Steering Committee*

Ron Anderson, Tjeerd Plomp (Chair), Richard Wolf, Ryo Watanabe

#### *Sampling Referee*

Colm O'Muircheartaigh

#### *Data Management*

Rien Steen

### Education systems:

Austria

Bulgaria\*

Germany

Greece

India

Israel

Japan

Latvia\*

Netherlands

Slovenia

Thailand\*

United States of America

Notes: \* = no stage 1 data collected.

### Populations

Students in primary education (grade 5), lower secondary education (grade 8) and upper secondary education (penultimate grade). Also, school-level respondents (school principal and technology coordinator) as well as teachers (computer education and existing subjects) from schools attended by the sampled students, were involved.

### Instruments

- School Questionnaire (principal part)
- School Questionnaire (technical part)

- Teacher Questionnaire Computer Education
  - Teacher Questionnaire Existing Subjects (international option)
- For collecting student data, the following tests/questionnaires were used:*
- Functional Information Technology Test (FITT)
  - Elementary programming test (national option)
  - Performance tests (internat. option): Word Processing
  - Attitudes scales:
    - Enjoyment
    - Relevance
    - Parental Support
  - Student description of computer use at school (general and within subjects) and home
  - Student background characteristics (age, gender, SES, etc.)



## **IEA-study: Second Information Technology in Education Study, Module I**

**Abbreviation:** SITES-M1

**Start:** 1997

**End:** 2000

**Year of Data Collection:** 1999

**Main study report:** Pelgrum, W.J. & Anderson, R.E. (Eds.). (1999). *ICT and the Emerging Paradigm for Life-Long Learning*. Amsterdam: IEA.

### **Staff**

#### *International Coordination Center*

Willem J. Pelgrum (International Coordinator), Klaas Bos, Joke Voogt

#### *International Steering Committee*

Ron Anderson (Chair), Chris Dede, Nancy Law, Fred-Arne Oedegaard, Jana Strakova, Ryo Watanabe

#### *Sampling Referee*

Colm O'Muircheartaigh

#### *Data Management*

Rien Steen

### **Education systems:**

Belgium-French	France	Norway
Bulgaria	Hungary	New Zealand
Canada	Iceland	Russian Federation
China- Hong Kong	Israel	Singapore
Chinese Taipei	Italy	Slovenia
Cyprus	Latvia	Slovak Republic
Czech Republic	Lithuania	South Africa
Denmark	Luxembourg	Thailand
Finland		

### **Populations**

School principals and technology coordinators from elementary, lower secondary and upper secondary schools.

### **Instruments**

Questionnaires for school principals and technology coordinators covering the following topics: Curriculum, ICT-Infrastructure, Staff development, Management and organization, Innovative practices, and Background information.

## **IEA-study: Second Information Technology in Education Study, Module II**

**Abbreviation:** SITES-M2

**Start:** 1999

**End:** 2002

**Year of Data Collection:** 2001

**Main study report:** Kozma, R.B. (Ed.). (2003). *Technology, innovation and educational change: a global perspective*. Eugene: ISTE.

### **Staff**

#### *International Coordination Center*

Robert B. Kozma (coordinator), Richard M. Jones, Raymond McGhee, Willem J. Pelgrum, Ronald D. Owston, Joke M. Voogt

#### *International Steering Committee*

Ron Anderson (Chair), Chris Dede, Nancy Law, Tjeerd Plomp, Jan Peter Stromsheim, Ryo Watanabe

### **Education systems:**

Australia	Germany	Philippines
Canada	Israel	Portugal
Chile	Italy	Russian Federation
China- Hong Kong	Japan	Singapore
Chinese Taipei	Korea	Slovak Republic
Czech Republic	Latvia	South Africa
Denmark	Lithuania	Spain, Catalonia
Finland	Netherlands	Thailand
France	Norway	United States

### **Populations**

Schools using innovative pedagogical practices. No national representative sampling.

### **Instruments**

Case study

## **IEA-study: Second Information Technology in Education Study, 2006**

**Abbreviation:** SITES-2006

**Start:** 2004

**End:** 2007

**Year of Data Collection:** 2006

**Main study report:** Law, N., Pelgrum, W.J. & Plomp, T. (Eds.). (2008). *Pedagogical practices and ICT use around the world: findings from an international comparative study*. CERC Studies in Comparative Education. Hong Kong: Comparative Education Research Centre, The University of Hong Kong, and Dordrecht: Springer.

### **Staff**

#### *International Coordination Center*

Tjeerd Plomp (Study Director), Willem J. Pelgrum (International Coordinator), Nancy Law

#### *International Steering Committee*

Ron Anderson (Chair), Alfons ten Brummelhuis

#### *Sampling Referee*

Christian Monseur

#### *Data Management*

IEA Data Processing Centre (Ralph Carstens and colleagues)

### **Education systems:**

Alberta, Canada	France	Norway
Catalonia, Spain	Hong Kong, SAR	Ontario, Canada
Chile	Israel	Russian Federation
Chinese Taipei	Italy	Singapore
Denmark	Japan	Slovak Republic
Estonia	Lithuania	South Africa
Finland	Moscow, Russian Fed.	Thailand

### **Populations**

All schools enrolling students at grade 8 and the mathematics and science teachers teaching those grades.

### **Instruments**

Questionnaires for school principals and technology coordinators covering ICT-infrastructure, Pedagogical practices, Vision, Staff development, Support

and Management/organization. Questionnaires for teachers covering a large number of topics, such as curriculum and pedagogy, ICT-use, assessment practices, self-reported competences, staff development, and background information about target classes and teacher characteristics.



## CHAPTER 7

# The Growth and Development of the Preprimary Project

*Jeanne E. Montie*

This chapter describes the IEA Preprimary Project from its inception in the early 1980s through the publication of the final volume in 2006. The chapter begins with a description of how the project was conceived and approved by the IEA, followed by a section relating the challenges of instrument development and sampling. The third section is a description of the day-to-day activities required to coordinate and move the project forward. The author worked on the project in various roles for 16 years and this section is based on her memories and those of people with whom she worked. The fourth section documents efforts at dissemination of the findings and the final section discusses the impact of the research.

### Conception of the Study

The idea for a cross-national study of early childhood education and child care was a dream of Marcel Crahay as long ago as 1979. Dr Crahay, a Professor of Educational Research at the Laboratoire de pédagogie expérimentale at the University of Liège in Belgium when the Preprimary study began, had a vision to create “an inventory of the way children live in each country of the world,” according to a 1989 interview in *Childhood Education* (Bridgman, 1989, p. 157). Working together with the IEA and early childhood experts Lilian Katz of the College of Education at the University of Illinois in the U.S. and Wolfgang Tietze of the Institute für Erziehungswissenschaft at Westfälische Wilhelms-Universität Münster, West Germany and others he developed a proposal for a far-reaching and unprecedented study of the lives of children before they enter primary school. The proposal

was first formally presented to the IEA at the General Assembly meeting in Toronto, Canada in 1982.

Crahay and his colleagues knew that to undertake such a study they would need the expertise and experience of the IEA. IEA experts had experience training teams of researchers within and across many countries and the IEA also had years of experience developing and using research instruments that were valid and reliable across countries. This would be a particular challenge in the Preprimary Project; early childhood was an area where few nationally valid instruments existed, let alone instruments that could be used across countries.

The researchers were aware that, because preprimary education was outside the mainstream of education in most countries, there would be challenges in developing and carrying out such a study that the IEA had not encountered in previous school-based studies. At the outset it was necessary to identify the settings in which preprimary children spend their time and to deal with the sometimes confusing terminology. In the United States alone, preprimary settings included, for example, day care centers, preschools, nursery schools, and family child care homes. In other countries preprimary settings were termed infant classes, crèches, child-minders, day nurseries, and play groups, among many others. Kindergarten, a term used in many countries, had almost as many different meanings. This variety of terms reflected the variety of care settings, including differences in the age-range served, length of the school day, sponsorship, region, and philosophy. There were also differences in the occupational titles of the adults who work with children in such settings, for example, teacher, child care worker, nursery nurse, sister, and child minder. These terms reflected the type of institution, as well as the training and educational background of these individuals.

In the following quote from the original proposal the authors summarize challenges to the study,

...pre-primary education differs in several important ways from those usually examined in the studies undertaken by IEA. In the

first place, it is less homogeneous and monolithic as a professional field both within and between countries than elementary or secondary education. Pre-primary education is in fact several different fields: child development, nursery education, social welfare, and so forth. Secondly, due to the variety of agencies responsible for institutions within it, there are great gaps in basic descriptive and utilization data with respect to it...Finally, in the absence of accepted and/or robust theories, the field is marked by sharp ideological/philosophical controversies... (Katz, Crahay, & Tietze, 1982, pp. 19-20)

At the 1983 annual meeting the General Assembly formalized the Preprimary Project and an International Steering Committee was appointed consisting of Lilian Katz as the chairperson, and Marcel Crahay, Wolfgang Tietze, and Richard Wolf (U.S.) as members. The HighScope Educational Research Foundation, located in Ypsilanti, Michigan, was designated as the International Coordinating Center (ICC) with HighScope president and noted early childhood researcher, David Weikart, as the International Coordinator. The final project proposal was approved at the General Assembly meeting in Singapore in 1984.

It was agreed that the study's central goal would be to examine different aspects of early childhood education, collectively called *quality of life* in the study, that enable children to do better in school by age 7, the age when children in all participating countries would have entered primary school. The study was to take place in three phases and was expected to run through 1994. The purpose of the first phase was to identify where preprimary children spend their time by collecting interviews with families of young children using household survey procedures based on nationally representative samples. In many countries this would provide new information about where preprimary children spend their time and how they spend their days. Once the preprimary care and education settings were identified in Phase 1, researchers would observe children and their caregivers in these settings in Phase 2 to determine what it is that children and their care-



givers actually do throughout a typical day. Phase 3 would be a follow-up study done after children received one year of formal schooling, looking for long-term effects of the preprimary experiences.

The study offered an unprecedented opportunity in many countries to gather information about services for preprimary children, the quality of services, how the services were used, and the impact of the children's preprimary experience on later skills and learning. In spite of the known difficulties, there was interest in participating in the study on the part of many IEA member countries--at least 17 in the beginning. Of course, before the study could get underway, funding had to be obtained and each country was responsible for obtaining funds for its own national project. Countries received funds from government agencies, universities and private foundations. Obtaining funds was always difficult and Crahay describes the process as "a thorn in the project's side...On the conceptual level, we were ready to start two years [before we actually began]" (Bridgman, 1989, p.157).

### **Challenges of Sampling and Instrument Development**

The design of the Preprimary Project was guided by a conceptual framework based on Bronfenbrenner's ecological model of development (1979, 1989). Ecological theory considers development to be the product of transactions between a child and his or her immediate environment, which is itself the product of forces operating in the broader social environment surrounding the family. Thus, the environment is conceptualized to have several levels, all of which directly or indirectly affect the developing child. Marcel Crahay developed the conceptual framework based on Bronfenbrenner's model (Crahay, 1990). The framework identified five groups of variables affecting children's development:

1. Family characteristics - household composition, parental education and occupation, parental beliefs about the relative importance of various areas of development.

2. Setting characteristics - teacher education and experience, equipment and materials, group size and staff-child ratio.
3. Teacher characteristics - beliefs about the relative importance of various areas of development, management of children's time, types of groupings, behaviors and interaction with children.
4. Children's activities - activities engaged in at the setting, involvement with other children and adults.
5. Children's developmental status - cognitive, language, and social development.

### **Decision Making and Communication**

The daunting task of the Steering Committee, ICC, and National Research Coordinators (NRCs) was to develop the methods and instruments to measure these variables in an impartial way across countries and cultures. To achieve this, the ICC made a major commitment to include all NRCs and their staff members in instrument development. According to Weikart, "from the outset the project was viewed as a group undertaking and the group firmly believed that the project must respect different cultures, range of economic development, historic language issues, ethnic relationships with countries, and similar fundamental issues" (Weikart, Olmsted & Montie, 2003, p. 6). Although all pledged cooperation toward a common goal, there were inevitable conflicts. As Weikart described the process,

The main problem was that a few NRCs, who were familiar with a more directive style and felt very knowledgeable in child development theory, wished to take control themselves of project design and instrument development rather than act in cooperation with all participating NRCs. Because they believed that they had the right scientific methods, they felt that they, or the ICC with their help, should select a theory of child development and create tests for specific hypotheses relating to that theo-

ry...Other NRCs were affronted by such a proposal, feeling it to be motivated by intellectual arrogance and a lack of sensitivity to the project's study of "their" children. (Weikart, Olmsted & Montie, 2003, p. 6)

The differences of opinion were not easily resolved; however, with help from the General Assembly, a system was agreed upon whereby at annual project meetings with NRCs, debate was held in check by limiting the number of times each NRC could comment on each issue. After three comments were made, the NRC was not recognized again until the next issue came to the table. Decisions could then be made by majority vote; unfortunately the contention resulted in the departure of two countries from the study after Phase 1.

Another challenge was communication with countries as the study progressed. In the early years of the study, most of the communication to NRCs relied on mailings. When the study began, in the mid-1980s, mail delivery to some countries was not very dependable, fax was just beginning to be widely used, and email was unheard of. Sending materials out for review and receiving feedback was an arduous process that took weeks, and sometimes, months. Annual planning meetings were held for NRCs and their staff; however, it could be difficult for some NRCs to obtain funding for travel. The meetings were most often held at the HighScope Retreat and Meeting Center in Clinton, Michigan, although occasionally NRCs were able to obtain funds to host a meeting and over the years the planning group had the opportunity to meet in Belgium, Greece, Italy, Spain, and Poland.

In addition to the practical problems of communication, there was the challenge of communicating clearly with all team members, most of whom were not native English speakers. As Weikart described processes used to make discussions as clear as possible, each country director had an opportunity to chair different sections of the annual meeting, around-the-table discussions allowed each director to summarize what he or she understood, and small group work sessions enabled personal reflection and discussion (Weikart, Olmsted & Montie, 2003).

Within this commitment to cooperation and open communication the work of instrument development and planning took place. An early decision was made to review each country's public policies regarding child care and education. NRCs from each country participating in Phase 1 wrote profiles on the status of preprimary education and care services in their country. This resulted in the first of the six books published about the Preprimary Project, *How Nations Serve Young Children* (Olmsted & Weikart, 1989), and laid the groundwork for the Phase 1 sample survey of households to record what type of care families use for their young children.

### **Designing the Sampling Plans**

In 1987 the project was fortunate to secure the services of Leslie Kish as international sampling referee. Dr Kish was a professor emeritus at the University of Michigan where he was a founder of the Institute for Social Research and author of the classic text, *Survey Sampling* (1965). Originally from eastern Europe, Dr Kish spent a great deal of time traveling and working throughout the world and was well known internationally. In his time abroad he loved to immerse himself in local culture, sampling local food and occasionally adopting native dress. He relished the challenge of working with NRCs from a variety of countries and developing representative national samples.

It was indeed a challenge to develop individual sampling plans for each participating country. Some NRCs had little or no sampling experience and few resources. The samples would have to be limited (400 - 1,000 children); thus, one challenge would come from spreading these small samples over large populations of households. Another challenge would be finding households with 4-year-old children, only 4 to 10 percent of households, so researchers would have to screen large numbers of families to come up with the desired sample. Throughout a period that spanned many months, Dr Kish and the ICC worked with all the NRCs; they educated them in general sampling issues,

helped them complete a lengthy sampling questionnaire and, ultimately, developed a unique sampling plan for each country.

### **Instrument Development**

While the sampling plans were being finalized, NRCs were also working together to develop instruments to use in the study. The Parent/Guardian Interview used in Phase 1 gradually evolved with reviews from experts and multiple field trials. During 1988 and 1989 the survey data were collected and sent to the ICC for analysis. At the same time, planning was well underway for Phase 2 of the Preprimary Project.

The unique task at the outset of Phase 2 was to develop instruments to measure the five categories of variables listed previously. Much of the information on family and setting characteristics and teacher background could be collected with questionnaires, but documenting what actually happened in early childhood settings was more problematic. The research team grappled with the question of how meaningful information could be collected across nations with a multitude of cultures, languages, educational systems, economic levels, and expectations for children. It was necessary to decide between two approaches. The first was to select a well-developed, theoretically based approach to early childhood care and education and measure the extent to which different national settings achieved these standards. The second was that the project would develop a nonjudgmental, descriptive observational system that would be based on teacher and child behavior without any assumption as to what was good practice and what was not. After extended discussion, the research team endorsed the second approach. As one NRC put it, "You're not going to use your theoretical system to measure my children" (Montie, Xiang & Schweinhart, 2007, p. 2).

It is important to note that the researchers chose direct observation, a difficult and expensive procedure, rather than using rating scales. The team rejected the use of rating scales because of the value judgments inherent in them. For example, the scoring of one widely used rating

scale depended extensively on ratings of the physical aspects of the setting; this would not be suitable for use in Nigeria, for example, where some early childhood programs took place outside under a tree. The group preferred to collect data that showed exactly what was happening in the class. A three-part time sampling method was designed that provided records of what teachers and children (separately) were actually doing on a minute-to-minute basis (Child Activities and Adult Behavior), while simultaneously recording what the teacher was asking the children to do (Management of Time). Each NRC provided videotapes from classrooms in their country and at the 1989 NRC meeting the group attempted to use the time sampling instruments while watching the videos. As Weikart reported, "...the NRCs were stunned as they watched the videotapes from the different countries. The 4-year-old children, the teachers, and the settings were obviously very different, yet so very similar...with minor adjustments their procedures could be used in all of the settings, and it would be possible to have a combined observational system..." (Weikart, Olmsted & Montie, 2003, p.9).

Over a period of several years the NRCs piloted the instruments in their countries and sent the data to the ICC for analysis. At the annual meetings discussions ensued, revisions were made and the instruments were sent out for further pilot testing. At the same time questionnaires and developmental status measures were drafted and tested. Finally, in the fall of 1991 the ICC distributed the final versions of the 12 instruments to be used in Phase 2 data collection.

## **Life at the International Coordinating Center**

### ***Personnel and Procedures***

As project teams in each of the participating countries learned about sampling and began to collect household survey data, an IEA research team was put in place at HighScope. Dave Weikart appointed Patricia Olmsted as Deputy International Coordinator in 1987 and she, in turn,

assembled her own work team. Over the years, team members came and went but two individuals, Shannon Lockhart and Jill Claxton, remained with the project from the late 1980s through the publication of the final volume in 2007. Jill and Shannon served many roles and grew with the project. They each came to the project with a background in education and, after years of being immersed in the research, both pursued further education in the research field. Initially, they assisted Pat Olmsted with the many and varied tasks that accompany coordination of a large project—everything from copying materials, sending out mailings, and maintaining orderly files, to hosting visitors at annual meetings. It soon became clear that they were capable of much more. Jill and Shannon took on major roles in instrument development, training, supervising data collection in the U.S., and maintaining communication with the NRCs. As the project moved on they learned data analysis skills and eventually each took on writing assignments for the various project publications.

After data were collected, each country with the capability provided its own country-based data analysis; however, all data were sent to the ICC for the cross-national analyses. In the early days it was sent on tapes and the HighScope team relied on personnel at the Institute for Social Research at the nearby University of Michigan (U of M) to provide data analysis. As the workload and funding increased, the project was able to hire its own data analysts who worked at HighScope. Over the years, expert analysis was provided by Zhenkui Ma, Mei-Yu Yu, Molly Gong, Alice Wan, Ted Jurkiewicz, Zongping Xiang and others. At first, the analyses were done using programs on the University of Michigan (U of M) server. It was necessary to login to the U of M computing center from HighScope, access SPSS, do the required analysis, and pick up the print-out at the computing center at U of M, a half hour away. It was always exciting when Molly or Mei-Yu arrived at HighScope with an enormous print-out of results or even raw data files to scrutinize and clean. The print-outs were stored in large binders and threatened to swallow up the small ICC offices, located on the top floor

of the old Victorian mansion that HighScope called home. Eventually, of course, the analysts were able to use software on their own personal computers to do their work and to store and share it online.

As mentioned previously, in the beginning most communication with the NRCs took place by mail. Pat Olmsted was a highly organized person and put in place an efficient method to track correspondence, copies of which filled banks of file cabinets in her office. Mailings were frequent and cumbersome. Data analysis print-outs, pilot test materials, drafts of chapters, and memoranda all had to be copied in multiples and mailed to addresses around the world. The availability of fax machines made communication easier, but there was more than one occasion when the NRC in Romania could not receive a fax because there was no fax paper available at her university. By the end of the project, email was the norm for communication to all of the NRCs, and the days of awkward mailings seemed like ancient history.

### *Annual Meetings*

It was always an exciting time when the annual NRC meetings took place at the HighScope Retreat and Meeting Center, a somewhat rustic, camp-like setting in the countryside where visitors shared bunkrooms and were treated to home-style American meals in the dining hall. The meetings were often held in November, close to the time when Americans celebrate their Thanksgiving holiday commemorating early American settlers giving thanks for the harvest. On one memorable occasion meeting participants from all over the world were treated to a traditional American Thanksgiving turkey dinner and pumpkin pie. Although the ICC staff and the NRCs and their accompanying staff worked hard and held meetings all day and into the evening, time was always set aside for fun and entertainment. Dave Weikart's wife, Phyllis, had an avid interest in traditional dance and she was always happy to spend an evening teaching folk dance from around the world to willing participants. Dave would often organize walks around the grounds after lunch. Occasionally it was



snowy when the meetings were held and for those NRCs from tropical climates, walking in snow and ice was a first. In fact, Jill Claxton and Shannon Lockhart put together a 'winter survival box' with extra boots, mittens, coats, and scarves for those who needed them. They were also enlisted to show the visitors around the local area and say the most requested destination was the local shopping mall.

In addition to hosting visitors at the annual meetings, over the years the ICC had the privilege to host several NRCs for longer visits. Mikko Ojala and Nóirín Hayes, NRCs from Finland and Ireland, respectively, collaborated on a joint project and each was able to spend an extended period of time working at the ICC. In 1995 Zongping Xiang, the NRC from China, came to the U.S. as a visiting scholar at the U of M and eventually came to work at HighScope permanently. Yemi Onibokun, from Nigeria, also spent time at HighScope doing writing and data analysis.

### *Managing the U.S. Preprimary Project*

At the same time that data analysis, instrument development, and report writing were taking place for the international project, staff at the ICC were responsible for carrying out the Preprimary Project in the United States. I, (Jeanne Montie), was hired as Coordinator for Phase 2 of the U.S. project in 1991 and was soon swept up in the activity. Sampling for the U.S. project was particularly problematic. Information gained from the Phase 1 survey provided a base for the Phase 2 sample and in many countries all that was necessary was to collect data from a random sample of classrooms and children in designated national and privately sponsored preprimary programs. In the U.S. there were few government-sponsored programs, a wide variety of privately sponsored programs, and many preprimary children spent their days at home or were cared for by relatives or families nearby. With the assistance of Leslie Kish, we developed a listing and screening sampling plan that required us to select census tracts at the six participating U.S. sites and proceed to knock on doors at each house or apartment in a

designated area and inquire as to whether there was a 4-year-old living there and, if so, enlist the family in the study. Jill Claxton was responsible for training the listing and screening teams from around the country and helping them iron out problems. One memorable difficulty occurred at the rural Mississippi site, where a flood had wiped out almost an entire town and the research team there had to travel miles and miles around cotton fields to find even one house.

After months of work, the samples at each site were established and the local research teams visited homes and schools to observe the children, teachers, and parents. The team in Los Angeles faced a particularly challenging situation when widespread rioting occurred after a jury acquitted the police officers accused of beating suspect, Rodney King, in 1992. The research team members had forged good relationships in the inner city community where they worked and proceeded to collect data through days of unrest in the city.

## **Dissemination Activities**

### *Videotapes from Around the World*

Dave Weikart and the other NRCs were so taken with the short videotapes of classrooms that each NRC brought to a planning meeting that he made the decision to send a team from HighScope to all of the participating countries to do more extensive videotaping in early childhood settings. HighScope would cover the travel costs while each host country would cover in-country costs. He envisioned editing the videos with commentary and making them available to teacher educators and policymakers internationally. They would provide a glimpse into other cultures and illustrate similarities and differences in practices in early childhood classrooms around the world.

Because of her experience in early childhood classrooms, Shannon Lockhart was selected to organize and lead the trips accompanied by HighScope videographer Kevin McDonnell. In three separate trips in

1992 and 1993 they traveled to a total of 12 Asian and European countries; in each country they videotaped a typical morning in one or two representative classrooms. Shannon has vivid memories of experiencing the bustle and crowds of Jakarta, rooftop play areas in Hong Kong skyscrapers, a student protest in South Korea, ancient ruins in Italy and Greece, and the reactions to new-found freedom in Poland and Romania. The videos were published by the HighScope Press (Lockhart, Weikart & Olmsted, 1994).

### *Writing Reports*

Because the Preprimary Project occurred in three phases, the researchers involved in it were usually working on several major tasks at any given time. As Phase 1 data were collected, instrument development and sample planning were underway for Phase 2. After the Phase 1 data were analyzed, it was time to begin writing the second volume, *Families Speak* (Olmsted & Weikart, 1994), while at the same time collecting data for Phase 2 of the study. The Phase 2 data collection began at the end of 1991 and was completed in most countries by the spring of 1992, as planned. However, many countries had funding difficulties that delayed the project and a complete data set was not available until 1994. The researchers did not consider this to be a problem as all tests were administered individually and there was no possibility of leaks of content of test items, nor was the content of the tests related to school curricula. The researchers did not believe that any cohort effects as a result of changes in country policy or economics would affect the results in the time period of data collection. The final sample included more than 5,000 children from more than 1,800 early childhood settings across the 15 participating countries.

The plan was to publish several volumes based on the Phase 2 findings and the research team was divided into small groups to plan each report. There continued to be a strong commitment to collaboration and each NRC contributed to the writing. This was an especially laborious process before email was widely available. Drafts of chapters

were sent to the ICC for editing and circulated among the NRCs for comment, all by mail. Although the NRCs were each responsible for disseminating country-based findings in their own countries, joint presentations were prepared for early childhood meetings in the U.S. and elsewhere. Often, last-minute changes were necessary when a presenting NRC could not get funding to travel.

In all, three volumes were published by the HighScope Press presenting the findings from Phase 2. *What Should Young Children Learn?* (Weikart, 1999) reports findings from interviews with teachers and parents regarding their beliefs about the relative importance of different areas of development for young children. *Early Childhood Settings in 15 Countries* (Olmsted & Montie, 2001) describes the management policies, patterns of operation, physical characteristics, and teacher background in preprimary settings. *A World of Preschool Experience* (Weikart, Olmsted & Montie, 2003) presents the findings from the direct observations in the preprimary settings.

Twelve of the 15 Phase 2 countries participated in Phase 3, which was carried out when the 4-year-old children who were observed and assessed in Phase 2, turned 7 years old (1995 -1997). The focus of data collection and analysis in Phase 3 was cognitive and language developmental status measures. The project achieved a remarkable median retention rate of 86% across countries (Montie, Xiang & Shweinhart, 2007), a testimony to the care that researchers took to track the children over time, a particularly difficult task in countries like the U.S. where families are highly mobile. With data collection complete in the late 1990s, staff at the ICC turned their attention to the cross-national longitudinal analysis.

As is inevitable in any project that spans more than two decades, staff changes and funding problems were the norm. However, most NRCs and a core staff at the ICC remained with the project for many years (see Table 1 for a complete list of participating countries and NRCs). As the year 2000 approached funding for the ICC was at a low point.

**Table 1.** IEA Preprimary Project Countries and Coordinators

Belgium	Coordinators: Arlette Delhaxhe, Geneviève Hindryckx
China	Coordinators: Shi Hui Zhong, ZongPing Xiang
Finland	Coordinator: Mikko Ojala
Germany	Coordinator: Hans-Günther Rossbach
Greece	Coordinators: Christos Frangos, Meni Tsitouridou
Hong Kong	Coordinator: Sylvia Opper
Indonesia	Coordinator: Diah Harianti
Ireland	Coordinator: Nóirín Hayes
Italy	Coordinators: Lucio Pusci, Egle Becchi
Nigeria	Coordinator: Olayemi Onibokun
Poland	Coordinator: Malgorzata Karwowska-Struczyk
Portugal	Coordinator: Joaquim Bairrão
Romania	Coordinators: Ioana Herseni, Viorel Nicolescu
Slovenia	Coordinators: Eva Bahovec, Marjan Setinc
Spain	Coordinators: Jesús Palacios, Mireia Montané, Delors Iduarte
Thailand	Coordinators: Nittaya Passornsiri, Pusadee Kutintara
United States	Coordinators: Helena Hoas, Jeanne Montie, Jill Claxton
International	Coordinating Center
HighScope	Educational Research Foundation
International	Coordinator: David P. Weikart
Deputy International	Coordinator: Patricia Olmsted

Dave Weikart retired from the presidency of the HighScope Foundation in 2000, but he remained in his post as International Coordinator for the Preprimary Project even while dealing with serious illness. Pat Olmsted took on extra duties, spending time on fund raising and traveling to international meetings during this period, and progress on the Phase 3 analysis slowed. Pat resigned from the project in 2002 when she was faced with illness herself. When Dave passed away in 2003, the project hit a low point in funding and morale; however, the remaining IEA team at HighScope, including Zongping

Xiang, Jill Claxton, Shannon Lockhart, and myself, was determined to finish the analysis and complete the final volume. With consultation and support from HighScope president Larry Schweinhart and funding from the HighScope Foundation and the Weikart Family Foundation, the final report was completed and presented at the General Assembly meeting in Taiwan in 2004. The small staff continued to work on the project as time and funding allowed and staff members were able to present the findings at professional meetings and publish the final volume as well as journal articles over the next several years.

### **Impact of the Project**

Of course it is not possible to know the full impact of the IEA Preprimary Project but some points deserve mention. At the time the study began, in the late 1980s, the types of preprimary education and care available around the world varied enormously according to country policy, local custom, educational philosophy and economic factors. Many educators and policy makers doubted that preprimary care and education had long-term effects on children's developmental outcomes; however, around the world the need for services for young children was growing as more and more mothers were entering the workplace.

In some of the countries participating in the study, little was known about how young children were cared for and regulations about the quality of such care were few or nonexistent. Thus, findings from Phase 1 of the study enabled educators and government officials in these countries to learn about existing services for young children and the need for quality care and education. Dramatic world events occurred as the study unfolded, in particular, the end of the Cold War and the dissolution of the Soviet Union. Several European countries participating in the study had lived under centralized communist rule and now had to find new ways forward. Findings from the Preprimary Project in

Poland, Romania, and Slovenia documented changes in preprimary care and education during this tumultuous period and provided data that were necessary for future planning and policymaking.

Even in countries with well-established child care and education systems, findings from the study, particularly Phase 2 findings that showed what children and teachers were actually doing in the classroom, provided the basis for reform. In Belgium, for example, Marcel Crahay and his staff determined that teachers in classrooms were using a directive educational approach and advocated for change to a more child-initiated approach. Mireia Montané, the NRC from Spain found that children in the preschool classes observed spent too much time waiting for direction from the teacher and was able to educate teachers accordingly. In China the research showed that the number of books in the home was related positively to child outcomes. This finding led educators to increase the number of books available in preschool settings. Other findings from the study in China led to educational reform allowing children more opportunities to choose their own activities. These examples are just a few of the many country-specific findings that informed child care and education in each of the participating countries.

The cross-national longitudinal findings have the potential for an even more profound impact on the future of care and education of young children. The Phase 3 cross-national results report the relationships between structure and process in children's age-4 settings and the children's cognitive and language development at age 7. The findings show that across the vast cultural, political, and economic differences in the countries that participated in the study, there were certain characteristics of early care and education settings that were related to better performance on age-7 cognitive and language assessments. For example, children from preprimary classrooms in which teachers allow them many opportunities to choose their own activities, performed better on age-7 language tests than children from classrooms in which teachers proposed a lot of preacademic activities. Children who spent less time in whole group activities, when all the children in

the class participated in the same activity, did better on cognitive tests at age 7 than those who spent more time in whole group activities. These and other findings from the Phase 3 study are discussed in detail in (Montie, Xiang & Schweinhart, 2007; Montie, Claxton & Lockhart, 2007; and Montie, Xiang & Schweinhart, 2006). The findings tell us that preprimary teaching practices do matter, how teachers set up their classrooms and the activities they propose for children make a difference. The findings highlight the importance of allowing children to be active participants in their own learning, to be able to choose their own activities, work individually or in small groups, and have ample materials available to them.

Of course, the study had impacts on the individuals who worked on it as well. Those of us at the ICC were stretched in ways we could not have predicted—learning new sampling, data analysis, and writing skills and making friends across the world. As staff from the ICC and many NRCs met to present the findings at meetings in Europe, Asia, and North America, new cultures were experienced and friendships deepened. A few individuals made radical life changes as a result of opportunities provided by their experiences with the Preprimary Project. Zongping Xiang is one of those. As a young educational researcher in China she was asked by the country's NRC to translate documents for the project into Chinese. She was sent to the University of Hong Kong to learn survey methodology and increased her knowledge of statistical procedures as she helped to analyze the data coming in. The publication of the Chinese findings from the Preprimary Project led to other opportunities in China to take leadership roles in new studies. She traveled to the U.S. in 1995 to work at the U of M as a visiting scholar and also spent time at HighScope. Eventually she was able to move to the U.S. and take employment at HighScope where she has served as the head of data analysis for the Preprimary Project and many other studies.

With the completion of the project those of us who worked on it have moved on to other pursuits with the satisfaction of knowing that we



helped to make a significant contribution to the knowledge base in early childhood education and care around the world.

## References

- Bridgman, A. (1989). The IEA Preprimary Project How a dream became a reality. *Childhood Education*, Spring, 157-159.
- Bronfenbrenner, U. (1979). *The ecology of human behavior*. Cambridge, MA: Harvard University Press.
- Bronfenbrenner, U. (1989). Ecological systems theory. *Annals of Child Development* 6, 187-249.
- Crahay, M. (1990). *Which research paradigm for the IEA longitudinal quality-of-life study*. Unpublished manuscript, University of Liège, Belgium.
- Katz, L. G., Crahay, M. & Tietze, W. (1982). The pre-primary education project. Proposal submitted to the IEA General Assembly.
- Kish, L. (1965). *Survey Sampling*. New York: Wiley & Sons.
- Lockhart, S., Weikart, D. P., & Olmsted, P. P. (1994). *Sights and sounds: High/Scope international videotape series*. [Set of 15 videotapes]. Ypsilanti, MI: High/Scope Press.
- Montie, J. E., Xiang, Z., & Schweinhart, L. J. (Eds.) (2007). *The role of preschool experience in children's development: Longitudinal findings from 10 countries*. Ypsilanti, MI: High/Scope Press.
- Montie, J. E., Xiang, Z., & Schweinhart, L. J. (2006). Preschool experience in 10 countries: Cognitive and language performance at age 7. *Early Childhood Research Quarterly* 21, 313-331.
- Montie, J. E., Claxton, J., & Lockhart, S. (2007). A multinational study supports child-initiated learning: Using the findings in your classroom. *Young Children* 62, 22-26.
- Olmsted, P. P., & Weikart, D. P. (Eds.). (1989). *How nations serve young children: Profiles of child care and education in 11 countries*. Ypsilanti, MI: High/Scope Press.

- Olmsted, P. P., & Weikart, D. P. (Eds.). (1994). *Families speak: Early childhood care and education in 11 countries*. Ypsilanti, MI: High/Scope Press.
- Olmsted, P. P., & Montie, J. E. (Eds.). (2001). *Early childhood settings in 15 countries: What are their structural characteristics?* Ypsilanti, MI: High/Scope Press.
- Weikart, D. P. (Ed.). (1999). *What should young children learn? Teacher and parent views in 15 countries*. Ypsilanti, MI: High/Scope Press.
- Weikart, D. P., Olmsted, P. P., & Montie, J. (Eds.). (2003). *A world of preschool experience: Observations in 15 countries*. Ypsilanti, MI: High/Scope Press.



## CHAPTER 8

# The International Civic and Citizenship Education Study (ICCS)

*Wolfram Schulz, Julian Fraillon  
& John Ainley*

The IEA *International Civic and Citizenship Education Study* (ICCS) investigated the ways in which young people are prepared for, and consequently ready and able to undertake their roles as citizens. Consequently, it studied student knowledge and understanding of civics and citizenship as well as affective and behavioral aspects of civics and citizenship such as value beliefs, attitudes, intended behaviors and current activities related to civic and citizenship education. Contextual data from education systems and schools were analyzed to help explain variation in these outcome variables.

ICCS built on the previous IEA studies of civic education (Amadeo et. al., 2002; Schulz & Sibberns, 2004; Torney-Purta et. al., 2001). In 1971 the IEA Civic Education Study included a 47 item test for 14 year olds in nine countries (Torney, Oppenheim & Farnen, 1975). In 1999 the IEA CIVED study included a 38 item test for 14 year old students in 28 countries (Torney-Purta et. al., 2001) and a 42 item test for 17-18 year olds in 16 countries (Amadeo et. al., 2002). Another chapter in this volume provides a history of these studies from the late 1960s to 2005 (Torney-Purta & Schwille, 2011). In addition to building on these studies ICCS is a response to the challenge of educating young people in changed contexts of democracy and civic participation.

### Scope

Thirty-eight countries participated in ICCS: Austria, Belgium (Flemish), Bulgaria, Chile, Chinese Taipei, Colombia, Cyprus, Czech Republic, Denmark, Dominican Republic, England, Estonia, Finland,

Greece, Guatemala, Hong Kong SAR, Indonesia, Ireland, Italy, Republic of Korea, Latvia, Liechtenstein, Lithuania, Luxembourg, Malta, Mexico, The Netherlands, New Zealand, Norway, Paraguay, Poland, Russian Federation, Slovak Republic, Slovenia, Spain, Sweden, Switzerland, and Thailand.

ICCS studied students in Grade 8 (on average including students who are approximately 14 years of age) provided that the average age of students in Grade 8 was 13.5 years or above.<sup>1</sup> Data were collected using a multi-stage cluster design where schools were sampled at the first stage and intact classrooms of students were sampled at the second stage. Within each school a teacher survey was conducted among a randomly selected sample of 15 teachers in each sampled school teaching regular school subjects to the students in the target grade.

In 2007 ICCS conducted a field trial. During 2009 ICCS surveyed more than 140,000 Grade 8 students collecting measures of knowledge and understanding of civics and citizenship affective and behavioral responses related to this learning area, as well as family and social background data. It also gathered data from over 60,000 teachers and principals in more than 5,000 schools in which those students were enrolled.

### **The Assessment Framework**

The assessment framework provided a conceptual underpinning for the international instrumentation for ICCS and a point of reference for the development of regional instruments (Schulz, Fraillon, Ainley, Losito, & Kerr, 2008). The assessment framework consisted of two parts. The *civics and citizenship framework* outlined the outcome measures addressed through the cognitive test and the student perceptions questionnaire. The *contextual framework* mapped the context factors expected to influence outcomes and explain their variation.

---

1. Where the average age of students in grade 8 was less than 13.5 years, England and Malta, students in Grade 9 was the defined population.

### ***The Civics and Citizenship Framework***

The ICCS *Civics and Citizenship Framework* underpinned the collection of student outcomes data and was organized around three dimensions: a content dimension specifying the subject matter to be assessed with test and questionnaire items within civics and citizenship; an affective-behavioral dimension that described the types of student perceptions and activities that were to be measured with questionnaire items; and a cognitive dimension that described the thinking processes to be assessed with test items.

#### **Civics and citizenship content domains**

The framework specified four content domains. The first content domain, *civic society and systems*, was concerned with the mechanisms, systems, and organizations that underpin societies. The second domain, *civic principles*, referred to the shared ethical foundations of civic societies. Civic participation dealt with the nature of the processes and practices that define and mediate the participation of citizens in their civic communities (often referred to as active citizenship). The framework recognized the centrality of the individual citizen through the *civic identities* domain. This domain refers to the personal sense an individual has of being an agent of civic action with connections to multiple communities. Together, these four domains encapsulated the civic and citizenship content to be assessed in ICCS. All four content domains were addressed from an affective-behavioral perspective and from a cognitive perspective.

#### **Civics and citizenship affective-behavioral domains**

Data relating to the affective-behavioral domains were collected by means of a questionnaire completed by students. Four affective-behavioral domains were specified:

- *Value beliefs* were defined as beliefs about the worth of concepts, institutions, people and ideas. These value beliefs help individuals resolve contradictions, and they form the basis of

how people see themselves and others. Value systems were envisaged as sets of value beliefs that individuals adopt and that, in turn, influence both attitudes and behavior.

- *Attitudes* were defined as states of mind or feelings about ideas, persons, objects, events, situations, and/or relationships. In comparison with value beliefs, attitudes were considered narrower in nature, more prone to change over time, and less deeply rooted. The attitudes relevant with respect to civics and citizenship included students' self-beliefs related to civics and citizenship; students' attitudes towards rights and responsibilities; and students' attitudes towards institutions.
- *Behavioral intentions* referred to student expectations of future action, not actual behavior. This affective-behavioral domain, assessed in the student perceptions questionnaire, required items that ask students about their intentions for civic action in the near future or as adults.
- *Civic-related behavior* is limited for 14-year-old students, and many activities for citizens are not available at this age. However, several civic-related behaviors can occur among 14-year-olds, and the aim was to capture these through the student background questionnaire.

### **Civics and citizenship cognitive domains**

In responding to the ICCS cognitive test items, students needed to know a core set of civic and citizenship content, apply cognitive processing to their civic and citizenship knowledge and relate their knowledge and understandings to real-world civic action. Two cognitive domains made up the cognitive processes that students were expected to demonstrate in the ICCS cognitive test. Data derived from the test items were used to construct a global scale of civic and citizenship knowledge and understandings of the four content domains and two content domains: The first cognitive domain, knowing, spec-

ified the civic and citizenship information that students are required to know. The second domain, reasoning and analyzing, detailed the cognitive processes that students required to reach conclusions broader than any single piece of knowledge, including the processes involved in understanding complex sets of factors influencing civic actions and planning for and evaluating strategic solutions and outcomes.

### *The Contextual Framework for ICCS*

ICCS investigated influences on civic-related learning outcomes and civic engagement. It took the view that young people develop their understandings about their roles as citizens in contemporary societies through a number of activities and experiences that take place within the contexts of home, school, classrooms, and the wider community. It recognized that young people's knowledge, competencies, dispositions, and self-beliefs are influenced by factors that can be envisaged at different levels in a multi-level structure. An individual student is located within overlapping contexts of school and home. Both contexts form part of the local community that, in turn, is embedded in the wider sub-national, national, and international context. The contextual framework for ICCS distinguished the following levels:

- *The wider community*: This level comprised the wider context within which schools and home environments work. Influences could operate at local, regional, and national levels. For some countries, the supra-national level might also be relevant as, for example, in member countries of the European Union.
- *Schools and classrooms*: This level of the contextual framework included factors related to the instruction students received, the school culture, and the general school environment. However, because of the sampling design for ICCS, school level and classroom level cannot be disentangled. Generally, only one classroom was selected within each sampled school.
- *Home environments*: This level of the framework comprised fac-

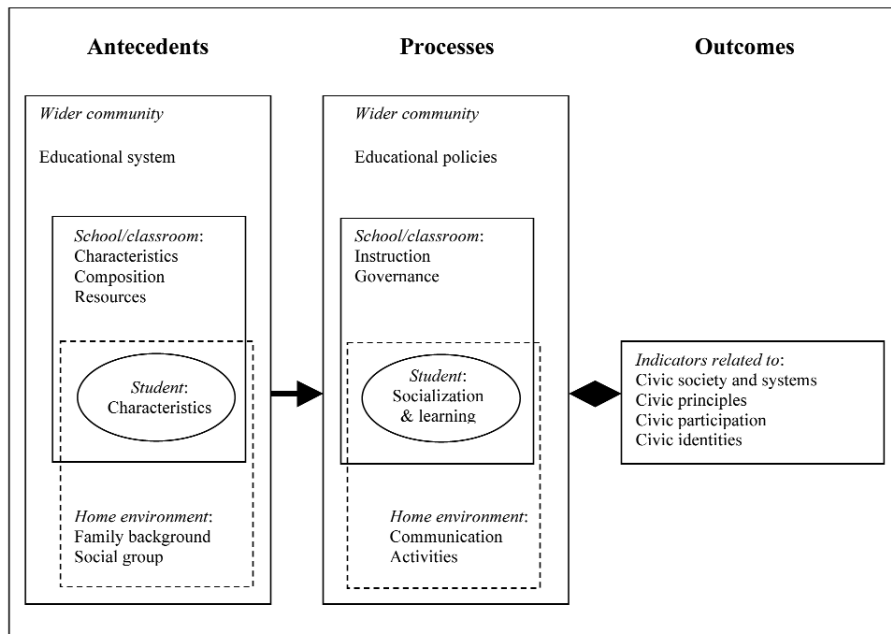


tors related to the home background and the social out-of-school environment of the student (for example, peer-group activities).

- *Individuals*: This level included the individual characteristics of the student.

Another important distinction in the context for civics and citizenship concerned whether those factors related to antecedents or processes:

- *Antecedents* were those pre-existing contextual factors that affected how student learning about and acquisition of, civic-related understandings and perceptions took place. For example, civic-related training of teachers may be affected by historical factors and/or policies implemented at the national level.
- *Processes* were those factors that characterize the ways in which civic-related learning and the acquisition of understandings,



**Figure 1.** Contexts for the development of learning outcomes related to civics and citizenship

competencies, and dispositions took place. They were considered to be constrained by antecedents and influenced by factors relating to the higher levels of the multi-level structure.

Antecedents and processes shape the outcomes at the level of the individual student. Learning outcomes related to civics and citizenship education at the student level also can be viewed as aggregates at higher levels (school, country) where they can affect factors related to process. For example, higher levels of civic understanding and engagement among students can influence the way schools teach civic and citizenship education.

Figure 1 illustrates which contextual factors might influence the learning outcomes of civic and citizenship education. The (double-headed) arrow between processes and outcomes signals a reciprocal relationship. It is important to emphasize that “feedback” occurs between civic-related learning outcomes and processes. For example, students with higher levels of civic knowledge and engagement are those students more likely to participate in activities (at school, at home, and within the community) that promote these outcomes. The (single-headed) arrow between antecedents and processes describes the relationship between these two types of factors at each level as uni-directional. However, higher-level processes can influence antecedents, and it is likely that, from a long-term perspective, outcomes may affect variables that are antecedents for learning processes.

## Data

The main survey data collection took place in the 38 participating countries between October 2008 and June 2009. The survey was carried out in countries with a Southern Hemisphere school calendar between October and December 2008, and in those with a Northern Hemisphere school calendar between February and May 2009. In a few countries, the teacher survey data collection was extended in order to achieve better participation rates.

### *International instruments*

ICCS developed and applied a number of instruments for the project in all countries. The international student test included 80 items (multiple-choice and open-ended response item types) in seven different clusters. These clusters were administered in a complete rotated design with seven randomly allocated booklets, each consisting of three 15-minute clusters. Therefore for each student the test was of 45 minutes duration. The ICCS test of Civic Knowledge included a link to the CIVED survey in 1999 through the use of a set of common items within the larger ICCS item pool. This link was used to assess changes in civic content knowledge in 17 countries.

A 40-minute international student questionnaire was used to obtain student perceptions about civics and citizenship as well as information about each student's background. The international student questionnaire was of 40 minutes duration and was administered after the international test booklets.

**Table 1.** Domain coverage in ICCS international student instruments

	<b>Content</b>				<b>Total</b>
	Civic society and systems	Civic principles	Civic participation	Civic identities	
<b>Cognitive</b>					
Knowing	15	3	1	0	19
Analyzing and reasoning	17	22	17	5	61
<b>Total</b>	<b>32</b>	<b>25</b>	<b>18</b>	<b>5</b>	<b>80</b>
<b>Affective-behavioral</b>					
Value beliefs	12	12	0	0	24
Attitudes	12	18	18	14	62
Behavioral intentions			21		21
Behaviors			14		14
<b>Total</b>	<b>24</b>	<b>30</b>	<b>53</b>	<b>14</b>	<b>121</b>

Table 1 shows the mapping of cognitive and affective-behavioral domains to content domains and the coverage of each cell with the cognitive and perceptions items included in the international student test and questionnaire. Cognitive items from both domains (*knowing; reasoning and analyzing*) and affective-behavioral items from two domains (*value beliefs and attitudes*) were developed for all four content domains. Items were not spread evenly across the cells in the table because there are differences in the relevance of the domains for Grade 8 students. Most items measuring the cognitive domain *knowing* were related to the content domain *civic society and systems* and affective-behavioral items measuring value beliefs were related only to two of the four content domains (*Civic society and systems, civic principles*).

### ***Regional instruments***

A set of regional instruments was also developed for ICCS. These focused on particular issues associated with civics and citizenship in three regions: Asia, Europe, and Latin America. The regional instruments were an innovative feature of ICCS. Their purpose was to allow assessment of region-specific aspects of civic and citizenship education. Participating countries in the regions of Asia, Europe, and Latin America could elect to participate in the relevant regional module. Nearly all of these countries decided to do so. Five countries participated in the Asian module, 24 in the European module, and six in the Latin American module. The regional instruments were administered after completion of the international student test and questionnaire. The Asian regional instrument was a questionnaire of 15 minutes duration. The European regional instrument consisted of a cognitive test (12 minutes) and a questionnaire (17 minutes). The Latin American regional instrument consisted of a cognitive test and region-specific questionnaire (each of 15 minutes).

The Asian regional questionnaire addressed issues of support for the preservation of traditional culture, obedience to authority, sense of Asian identity and attitudes to one's own country, as well as views on

the importance of morality and spirituality for being a good citizen.

The European regional instruments investigated students' civic knowledge in a European context as well as their attitudes, perceptions and behaviors in relation to European civic issues, institutions and policies: European citizenship and identity, intercultural relations in Europe, free movement of citizens in Europe, European policies, institutions and participation and European language learning.

The Latin American regional instruments investigated region-specific aspects of civic knowledge, perceptions of public institutions, forms of government, corrupt practices as well as obedience to the law. They also formed the basis for reporting on dispositions of student towards peaceful coexistence including attitudes toward their country and the Latin American region, sense of empathy, tolerance toward minorities as well as attitudes toward use of violence.

In addition to the international and regional instruments, ICCS offered several international options in the questionnaires. These options concerned students' ethnicity, household composition, and religion. Nineteen national centers chose to include the item on ethnicity, 37 national centers opted to include the item on household composition, and 28 chose to include the items on religion in the student questionnaire. Three national centers opted for asking only some of the items on students' religion.

### *Teacher, school and national contexts questionnaires*

ICCS also included a set of instruments designed to gather information from and about teachers, schools, and education systems. There was a 30-minute teacher questionnaire through which respondents provided their perceptions of civic and citizenship education in their schools and information about their schools' organization and culture as well as their own teaching assignments and backgrounds. In addition, ICCS included a 30-minute school questionnaire through which principals provided information about school characteristics, school

culture and climate, and the provision of civic and citizenship education in the school.

A national contexts survey was administered online and collected information about national contexts for civic and citizenship education. National research coordinators (NRCs) coordinated the information procured from national experts. This information concerned the structure of the education system, civic and citizenship education in the national curricula, and recent developments in civic and citizenship education.

### Outcome Measures

ICCS reported outcomes of civic and citizenship education on a number of international scales derived from the student test and the student questionnaire. This section describes those scales. The next section provides an overview of the reports that have been generated from ICCS.

The cognitive test items were scaled to obtain scores of civic *knowledge and understanding*. Civic knowledge was measured on a scale where the international average was set to 500 scale points, with a standard deviation of 100 scale points. ICCS revealed considerable variation across and within countries in the extent of civic knowledge. About half of the variation was recorded at the student level, about a quarter at the school level, and a further quarter across countries. The average civic knowledge scores for countries ranged from 380 to 576—a range equivalent to almost two international student-level standard deviations. The difference between the bottom quartile and the top quartile (i.e., covering the middle half of the averages for countries) was about 60 scale points. There was even greater variation in civic knowledge within the participating countries. For example, the distance between the lowest 5 percent and the highest 95 percent of civic knowledge scores was almost equal to 300 scale points. There were quite substan-

tial differences across countries in the within-country variation as well as in the extent to which this variation was associated with differences among schools.

The civic knowledge scale reflects progression from being able to deal with concrete, familiar, and mechanistic elements of civics and citizenship through to understanding the wider policy climate and institutional processes that determine the shape of civic communities. Analysis of the student achievement data led to the establishment of three proficiency levels. The lowest level was characterized by engagement with the fundamental principles and broad concepts that underpin civic and citizenship and by a mechanistic working knowledge of the operation of civic, civil, and political institutions. The second level was characterized by knowledge and understanding of the main civic and citizenship institutions, systems, and concepts as well as an understanding of the interconnectedness of civic and civil institutions and relevant operational processes. The top level was characterized by the application of knowledge and understanding to evaluate or justify policies, practices, and behaviors based on students' understanding of civics and citizenship.

The student questionnaire items were used to generate scales reflecting student value beliefs: support for democratic value beliefs, support for the importance of conventional citizenship and support for the importance of social-movement related citizenship. They also generated a number of scales reflecting attitudes: interest in political and social issues, civic self-concept, support for equal gender rights and responsibilities; support for equal rights and responsibilities for ethnic/racial groups; support for equal rights and responsibilities for immigrants; confidence in schools participation; trust in institutions; citizenship self-efficacy and attitudes toward country.

The student survey items were used to generate measures of behavioral intentions and behaviors such as: expectations to participate in legal protest activities; expectations to participate in illegal protest

activities; expectations to participate in elections; expectations to engage in active political participation; expectations to participate in informal political participation; reports on civic participation outside of school; and reports on civic participation at school.

## Reports

A number of reports have been, or are being, published from ICCS. International findings have been published in a Initial Findings report (Schulz, Ainley, Fraillon, Kerr & Losito, 2010a) and an International Report (Schulz, Ainley, Fraillon, Kerr & Losito, 2010b). Regional reports have been produced for Europe (Kerr, Sturman, Schulz & Burge, 2010) and Latin America (Schulz, Ainley, Friedman & Lietz, 2011). A regional report for the Asian region is to be released shortly. Furthermore, there will be an ICCS Encyclopedia with national chapters about civic and citizenship education in participating countries. The Technical Report for ICCS will describe details on instrument development, data collection procedures and analysis. In addition, many national centers have elaborated national reports focusing on the results for their educational systems in comparison with other countries.

## References

- Amadeo, J., Torney-Purta, J., Lehmann, R., Husfeldt, V., & Nikolova, R. (2002). *Civic knowledge and engagement: An IEA study of upper secondary students in sixteen countries*. Amsterdam: International Association for the Evaluation of Educational Achievement (IEA).
- Schulz, W., & Sibberns H. (Eds.). (2004). *IEA Civic Education Study*. Technical Report. Amsterdam: IEA.
- Schulz, W., Ainley, J., Fraillon, J., Kerr, D., & Losito, B. (2010a). *Initial Findings from the IEA International Civic and Citizenship Education Study*. Amsterdam: International Association for the Evaluation of



Educational Achievement (IEA).

Schulz, W., Ainley, J., Fraillon, J., Kerr, D. & Losito, B. (2010b). *ICCS 2009 International Report. Civic Knowledge, Attitudes and Engagement among Lower Secondary School Students in 38 Countries*. Amsterdam: IEA.

Schulz, W., Fraillon, J., Ainley, J., Losito, B., & Kerr, D. (2008) *International Civic and Citizenship Education Study*. Assessment framework. Amsterdam: International Association for the Evaluation of Educational Achievement (IEA).

Torney-Purta, J., Lehmann, R., Oswald, H., & Schulz, W. (2001). *Citizenship and education in twenty-eight countries*. Amsterdam: International Association for the Evaluation of Educational Achievement (IEA).

## CHAPTER 9

# IEA International Research Conferences (IRC) from 2004 to 2008

*Constantinos Papanastasiou  
& Elena C. Papanastasiou*

### Origins of the IRC Conference

The purpose of this chapter is to explain the concept and rationale behind the IEA International Research Conferences (IRC); to describe the three IRC conferences that have taken place so far; and to highlight the ways in which we see the IRC conferences heading in the future.

Although during the first 40 years of IEA, an overwhelming amount of data had been collected, there were no conferences nor any type of platform specifically designed to present and share IEA study findings. Constantinos Papanastasiou, the Cyprus National Research Coordinator (NRC) to the IEA, however, noted that often a group of NRCs would collaborate to organize symposiums or paper presentations for just such a purpose. In light of this, Constantinos suggested to various NRCs that perhaps IEA should organize an international conference in which secondary analyses results of the IEA projects could be shared. Most delegates, almost 80%, were positive. So, although a few members expressed reservations since there were already so many conferences worldwide, because the general conclusion was positive, Constantinos decided to make an oral proposal to IEA's Standing Committee about his vision and ideas for such a conference.

Thus in a Standing Committee meeting held in Amsterdam, Constantinos presented his proposal for an international conference. However, whether it was because the Standing Committee had more important issues to work on, or because the proposal might not have been properly explained, no decision was taken regarding the conference.

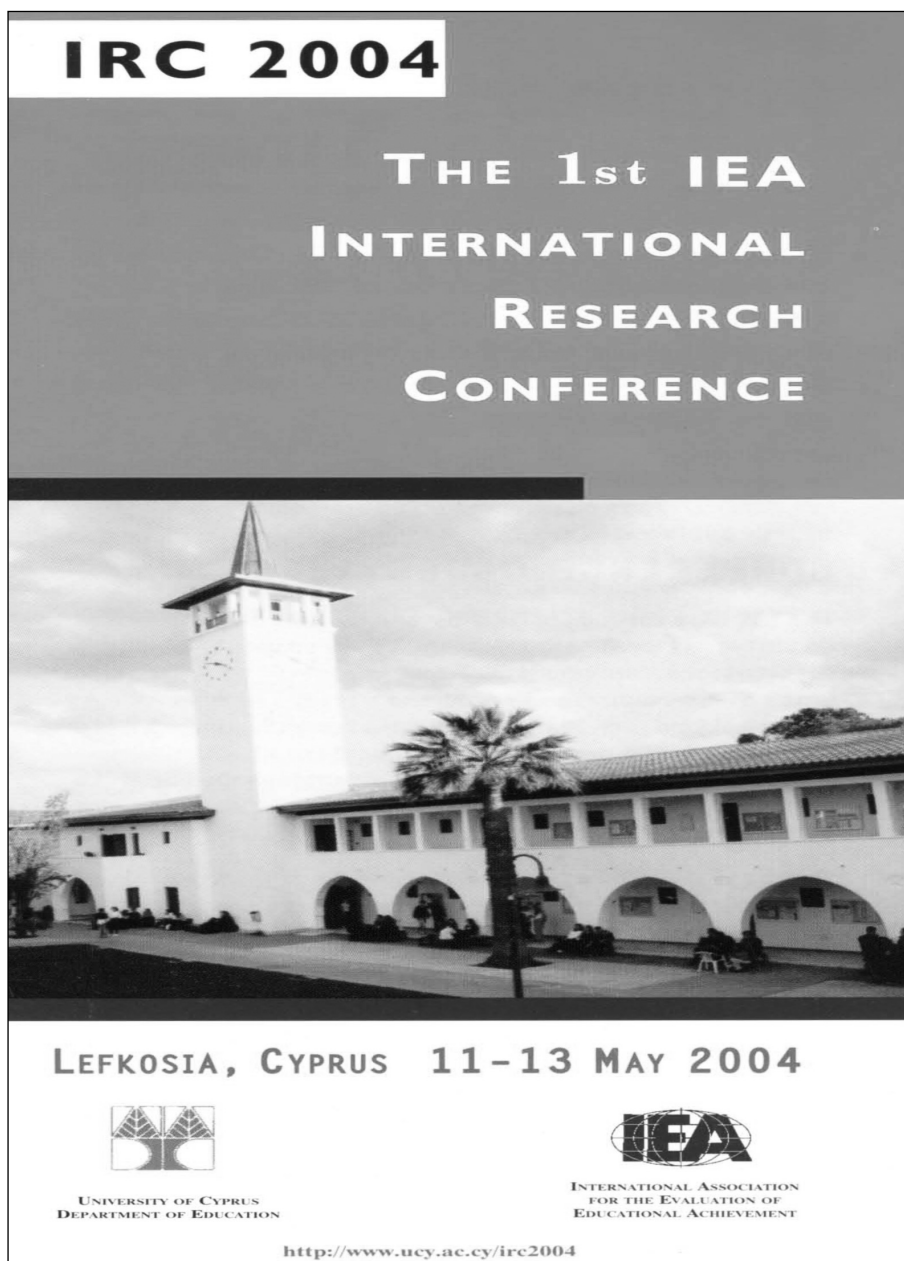
The following year, whilst participating in the General Assembly that took place in Thailand (in 2000), Constantinos discussed this issue with Dr Barbara Malak-Minkiewicz, Manager of Membership Relations of the IEA. Her advice was straightforward, "If you do not take the initiative to organize such a conference, no one will". This response was a strong incentive for Constantinos to proceed and to organize the 1st IEA International Research Conference (IRC).

Thus, in 2002, at the GA meeting in Marrakech (Morocco), an official invitation was extended to the GA Members to participate in the 1st IEA International Research Conference. This conference, called the **1ST IEA INTERNATIONAL RESEARCH CONFERENCE (IRC-2004)**, would be organized by the Department of Education of the University of Cyprus, and would be held at the University of Cyprus campus in Lefkosia (Nicosia), from Monday, the 11th of May to Wednesday, the 13th of May 2004.

### **First IEA International Research Conference (IRC-2004)**

The aim of the IRC-2004 Conference was to provide an international forum for the exchange of ideas and information on critical educational issues in education evolving from secondary analyses of data from IEA studies. With its objective to foster creative dialogue among scholars and researchers, the conference aimed at providing greater understanding of the numerous roles that education plays in the development of nations and in shaping individuals. Because of its international scope, the conference also aimed to examine issues in both a comparative and global context with the ultimate aim to enhance pedagogical knowledge and implement positive change. A final goal of this conference was to create a global network of researchers involved in secondary analyses of the IEA studies at the national, regional and global levels.

A flyer was created for the conference which was given to the GA members (Figure 1) and emailed to National Research Coordinators



**Figure 1.** The first page of the flyer which was given to the GA Members at the Morocco meeting during the official invitation to participate in the 1st IEA International Research Conference (IRC-2004)

involved in IEA studies and via them to researchers actively conducting secondary analysis of IEA data. The IRC logo (Figure 2), which was obviously included in the flyer, includes the words IEA circumscribed by the symbol of the earth, and denotes an international association of



Figure 2. IRC-2004 logo

prominent research institutions, universities, and ministry of education units of 58 countries around the world. The aims of IEA, as specified in its mission statement, are "[to undertake] the conduct of comparative studies focusing on educational policies and practices in order to enhance learning within and across systems of education." Based on the IEA logo, the IRC logo that was created included the words IRC circumscribed by the symbol of the earth, with the addition of the year that the conference would take place.

The call for papers was as follows:

You are invited to submit a proposal for the IRC-2004 conference, based on one of the following thematic strands: TIMSS, CIVED, SITES, PPP. Each proposal is intended to report SECONDARY ANALYSIS research results, and must be received no later than 1 September 2003.

**Criteria for acceptance**

- a. Topic: Originality, choice of problem. Importance of issues, relevance to program themes.
- b. Contribution to Education: Academic/educational importance, theoretical/practical significance.
- c. Analysis and Interpretations: Significance of conclusions; implications for research, practice policy, development of ideas; relationship of conclusions to findings; generalizability, or usefulness of findings or concepts.

The Scientific/Program Committee for IRC-2004 consisted of the following persons:

Constantinos Papanastasiou	Chair, University of Cyprus, Cyprus
Ray Adams	ACER, Australia
Albert Beaton	Boston College, USA
Eugene Gonzalez	Boston College, USA
Andreas Demetriou	University of Cyprus, Cyprus
Robert Kozma	SRI International, USA
Frederick Leung	University of Hong Kong, Hong Kong
Svein Lie	University of Oslo, Norway
George Marcoulides	California State University-Fullerton, USA
Michael Martin	Boston College, USA
Ina Mullis	Boston College, USA
Demetrios Natsopoulos	University of Cyprus, Cyprus
David Nevo	Tel Aviv University, Israel
Hans Pelgrum	University of Twente, The Netherlands
George Philippou	University of Cyprus, Cyprus
Tjeerd Plomp	University of Twente, The Netherlands
Jack Schwillie	Michigan State University, USA
Judith Torney-Purta	University of Maryland, USA
Hans Wagemaker	Executive Director of IEA

In addition to the overall aims of the IEA conference, the Local Organizing Committee had an additional single, but very important, objective for this conference. Since it would be the very first IEA conference, it was crucial that the organization of the conference be of a very high standard, as this would encourage, if not ensure, that this initiative would continue.

For this first conference, 84 proposals were submitted and the Scientific/Program Committee finally accepted 79% of these. The characteristics of the accepted proposals were as follows: TIMSS 52%, PIRLS 22%, CivEd/ICCS 17%, and SITES 9%.

The IRC-2004 took place on May 11-13, 2004. During the official opening four persons welcomed the participants, the Chair of the Organizing Committee Professor Constantinos Papanastasiou, the Chair of IEA Professor Alejandro Tiana, the Rector of the University of Cyprus Prof Stavros Zenios, and the Minister of Education and Culture Mr Pefkios Georgiades.

In the first two days of the conference, two keynote speakers addressed the participants in plenary sessions. The first speech, which is included in this book, was entitled, "IEA: International Studies, Impact and Transition", and was delivered by Hans Wagemaker, Executive Director of IEA (see Wagemaker, 2011). The second speech, which is also included in this book, entitled, "The contribution of IEA to Research and Education", was given by Richard M. Wolf, Teachers College, Columbia University, USA (see Wolf, 2011).

During the conference four volumes with the proceedings were distributed to the participants (Papanastasiou, 2004a, 2004b, 2004c, 2004d). The first two volumes included papers on TIMSS (Papanastasiou, 2004a, 2004b), Volume 3 included the papers on PIRLS (Papanastasiou, 2004c), whilst Volume 4 included papers on CivEd and SITES (Papanastasiou, 2004d).

Given the special character of the event, we have included the welcome speech for the IRC 2004, made by the Chair of IEA at the time, Professor Alejandro Tiana:

### **Welcome from the Chair of IEA, Professor Alejandro Tiana**

It is my pleasure to welcome you all to the International Research Conference (IRC-2004) organized by the International Association for the Evaluation of Educational Achievement (IEA) and the University of Cyprus. As you already know, this is the first conference we have organized with the specific objective of presenting and discussing research work based on the studies developed by IEA throughout its long history.

In the late 1950s IEA began to design and implement studies

aimed at evaluating educational outcomes from an international, comparative perspective. Forty years on, the number of projects that have been developed is really impressive, touching different curricular areas (reading, writing, mathematics, sciences) and exploring different educational fields (civic education, ICT in education, early childhood). A great number of researchers around the world have been involved in such studies and contributed to this collective task. It is my obligation to express to all of them our appreciation for the work done.

The research undertaken by the IEA throughout its history has made a significant contribution to the development of our education systems. By participating in the studies, countries and researchers have gained many and various benefits.

First of all, countries have received comparative information on the characteristics of their education systems, the scope of the curricula which they have both planned and implemented, and the results they have actually achieved. This information has enabled them to position themselves appropriately in an international context.

Secondly, this comparison has helped countries to identify the strengths and weaknesses of the education given in schools. They have not only learned whether their students are in a better or worse position on an international scale, but they have also received additional information to help explain those results.

Thirdly, countries have learned from the experience of others, comparing their respective situations and analyses. Learning from the satisfactory experiences of others is a good way to move ahead, and it has been encouraged by international meetings and publications resulting from IEA studies.

In addition to these contributions, participation in IEA studies has enabled countries and researchers to learn new performance assessment methods and procedures. Involvement in this type of transnational endeavor has assisted many national evaluation systems, contributing to the training of experts in the field and the improvement of study design and analysis.



It can be said that IEA has contributed to the setting-up and development of a worldwide community of researchers in educational evaluation. Now we are taking a step forward, allowing that community to share their interests, approaches, concerns and findings. Data and knowledge provided by IEA studies are so relevant and massive that they deserve further analyses. It is not enough to publish our project reports and disseminate our main results. We have the obligation to contribute to the improvement of our education systems. This conference will meet our expectations if it gives us the opportunity to present and discuss our research work. This is why I want to invite you all to participate in the best tradition of scientific rigor and intellectual openness. Welcome to IRC-2004!

After the conference the IEA Standing Committee decided that IEA would hold an international conference every two years. So we can safely conclude that the first IRC conference was a success, and that it managed to achieve all the goals that it had set out to achieve.

### **Second IEA International Research Conference (IRC-2006)**

The 2nd IEA International Research Conference was hosted by Tom Loveless of the Brookings Institute, Washington, DC, USA on November 9-11, 2006. The Brookings Institute, the IEA member institute for the USA, is a private nonprofit organization devoted to independent research and innovative policy solutions. For more than 90 years, Brookings has analyzed current and emerging issues and produced new ideas of importance for the nation and the world. Research at the Brookings Institute is conducted to inform the public debate, not to advance a political agenda. The Institute's goal is to provide high-quality analysis and recommendations for decision-makers in the USA and abroad.

The Members of the Local Organizing Committee of the 2nd IEA conference, who were all affiliated with the Brookings Institute were: Tom Loveless (Chair), Gladys L. Arrisueno, Katharyn Field-Mateer and

Lesley Pruitt. For this conference, 67 proposals were submitted, of which 18% were rejected. Of these accepted papers 55% were based on data from TIMSS, 27% from PIRLS, 13% from CivEd and 5% from SITES. Moreover, eight commissioned papers focusing on mathematics achievement through secondary analysis of IEA data were also presented at the conference. The authors of the chapters of the commissioned papers for this volume were Ina Mullis and Michael O. Martin, Jan-Eric Gustafsson, William Schmidt and Richard Houang, Jeremy Kilpatrick, Vilma Mesa and Finbarr Sloane, Laura Hamilton and Jose Felipe Martinez, Gabriela Schütz, Elena Papanastasiou, Dougal Hutchison and Ian Schagen.

All accepted papers were published in two volumes. The papers in Volume 1 of the proceedings have as their central focus the Trends in Mathematics and Science Study (TIMSS), and included 22 papers (IEA, 2007a). Volume 2 brings together 25 papers that focus on the Progress in International Reading Literacy Study (PIRLS), the Second Information on Technology in Education Study (SITES), and the Civic Education Study (CivEd) (IEA 2007b). However, a third volume edited by Tom Loveless was also produced, which included the eight papers that were commissioned by the Institute, in addition to an introductory chapter by the editor (Loveless, 2007).

**The Scientific/Program Committee for IRC-2006** consisted of the following persons:

Constantinos Papanastasiou	Cyprus (Chair)
Ronald Anderson	United States
Julius Bjornsson	Iceland
Robert Garden	New Zealand
Eugenio Gonzalez	United States
Jan-Eric Gustafsson	Sweden
Andris Kangro	Latvia
Kerry Kennedy	Hong Kong
Rainer Lehmann	Germany

Svein Lie	Norway
Pirjo Linnakyla	Finland
Christian Monseur	Belgium
Tjeerd Plomp	The Netherlands
Martine Remond	France
David Robitaille	Canada
Judith Torney-Purta	United States
Zacharias Zacharias	Cyprus
Ruth Zuzovsky	Israel

During the official opening of the IRC 2006, Dr Hans Wagemaker, the Executive Director of IEA, welcomed the participants as follows:

Dear Colleagues,

I would like to welcome you on behalf of the International Association for the Evaluation of Educational Achievement to our second International Research Conference.

While IEA is widely recognized for its work in conducting large-scale, comparative assessments of student achievement, it is equally committed to ensuring that the outcomes of these studies are widely disseminated, and that resulting analyses of the data that are collected impact the educational policy process. This conference is one of the mechanisms through which IEA attempts to encourage not only the further analysis and use of IEA data but also the science of large-scale assessment.

We hope that this meeting will be professionally rewarding for all of you and I look forward to meeting you in Washington.

### **Third IEA International Research Conference (IRC-2008)**

The 3rd IEA International Research Conference was hosted by the Science Education Center and the National Taiwan Normal University, Taipei, Chinese Taipei, on September 18-20, 2008. The conference was held at the Grand Hotel in Taipei City, while the conference was preceded by a 2-day training workshop on secondary data analysis (16-17 September).

**The scientific/paper review committee for the conference** consisted of the following persons:

Constantinos Papanastasiou	Cyprus (Chair)
Chun-Yen Chang	Chinese Taipei
Robert Garden	New Zealand
Eugenio Gonzalez	United States
Jan-Eric Gustafsson	Sweden
Ali Kiamanesh	Iran
Hwawei Ko	Chinese Taipei
Chen-Yung Lin	Chinese Taipei
Michael Martin	United States
Christian Monseur	Belgium
Ina Mullis	United States
Tjeerd Plomp	Netherlands
David Robitaille	Canada
David Rutkowski	Germany
Leslie Rutkowski	Germany
Wolfram Schulz	Australia
Knut Schwippert	Germany
Jouni Valijarvi	Finland
Ruth Zuzovsky	Israel

The local organizing committee was the following:

Tsung-Hau Jen	National Taiwan Normal University
Chen-Yung Lin	National Science Council
Chun-Yen Chang	National Taiwan Normal University
Che-Di Lee	National Taiwan Normal University
Chia-Li Chen	National Taiwan Normal University

The IRC-2008 conference, which had the same goals as the previous conferences, received 96 proposals, of which 86% were accepted for the conference. In addition to the quality indicators used for the selection of the proposals, two additional criteria were taken into account, which were that each proposal should:

1. be based on the IEA studies, TIMSS, CIVED, PIRLS, SITES and older IEA studies.
2. have reported secondary analysis results.

Overall, the characteristics of the accepted proposals in percentages are:

- 47% for TIMSS
- 33% for PIRLS
- 11% for SITES, and
- 9% for CivEd.

All of these papers, along with the papers of the other IRC conferences, can be found on the website of the IEA ([http://www.iea.nl/iea\\_irc.html](http://www.iea.nl/iea_irc.html)).

### **Some Final Remarks**

For all three conferences, the overall figures for the accepted proposals, in percentages are:

- 51% for TIMSS
- 28% for PIRLS
- 12% for CivEd and
- 9% for SITES

What is evident from these percentages is the clear preference of researchers to use TIMSS data since half of all the presentations were based on TIMSS. However, the PIRLS presentations are not negligible, especially if one considers the fact that PIRLS is based on one subject area while TIMSS is based on two subjects. What is a bit surprising, however, is the relatively low percentage of researchers using SITES data, especially because of the increasing importance that is placed on information technology in education.

Following up on the now-established IRC tradition, during the IRC-2008 conference in Taipei, the Chair of the Scientific/Program Committee met with the organizers for the IRC-2010 conference, namely Professor Jan-Eric Gustafsson (University of Gothenburg), and Professor Liv Gronmo (University of Oslo) to discuss some issues related to the upcoming conference. In this planning meeting, among

others, they came up with two basic goals:

1. To increase the number of presenters in the IRC-2010 conference and
2. To improve the quality of the papers.

The IRC-2010 conference did manage to achieve these goals, and it is our conviction that the IEA will continue to be successful in achieving and meeting its goals for future IEA conferences.

## References

- IEA (2007a). *The second IEA International Research Conference. Proceedings of the IRC 2006 (Vol. 1)*. New Zealand: Author.
- IEA (2007b). *The second IEA International Research Conference. Proceedings of the IRC 2006 (Vol. 2)*. New Zealand: Author.
- Loveless, T. (Ed.). (2007). *Lessons learned. What international assessments tell us about math achievement*. Washington DC: Brookings Institute Press.
- Papanastasiou, C. (Ed.). (2004a). *Proceedings of the IRC-2004:TIMSS (Vol. 1)*. Lefkosia: IEA.
- Papanastasiou, C. (Ed.). (2004b). *Proceedings of the IRC-2004:TIMSS (Vol. 2)*. Lefkosia: IEA.
- Papanastasiou, C. (Ed.). (2004c). *Proceedings of the IRC-2004:PIRLS (Vol. 3)*. Lefkosia: IEA.
- Papanastasiou, C. (Ed.). (2004d). *Proceedings of the IRC-2004:CivEd-SITES (Vol. 4)*. Lefkosia: IEA.
- Wagemaker, H. (2011). IEA: International studies, impact and transition. In C. Papanastasiou, T. Plomp, & E. Papanastasiou (Eds.), *IEA 1958-2008: 50 years of experiences and memories*. Nicosia: Cultural Center of the Kykkos Monastery.
- Wolf, R. M. (2011). The contribution of IEA to research and education. In C. Papanastasiou, T. Plomp, & E. C. Papanastasiou (Eds.), *IEA 1958-2008: 50 years of experiences and memories*. Nicosia: Cultural Center of the Kykkos Monastery.





PART B

**THE CONTRIBUTION OF IEA  
TO EDUCATIONAL RESEARCH**





## CHAPTER 10

# **The Relationship of IEA to Some Developments in Educational Research Methodology and Measurement during the Years from 1962 to 1992**

*John P. Keeves & Petra Lietz*

The marked expansion of education at all levels from early childhood education to primary and secondary schooling, as well as to universities and institutes of technology and to programs of lifelong learning and recurrent education in the workplace, led to the development of educational research as a major field of scholarly inquiry during the second half of the twentieth century. Research into education involves many disciplines in addition to psychology and operates with large numbers of individuals, institutions and systems at several levels across the approximately 200 states of the world. Moreover, it necessarily involves change over time since it is specifically concerned with the learning and development of young people, who are being taught in an increasingly complex world. During the years from 1962 to 1992 the authors of this chapter were actively engaged in the research studies of the International Association for the Evaluation of Educational Achievement (IEA). IEA formed part of this movement for the advancement of education and educational research. This chapter examines the links between IEA and the developments that occurred in educational research methods and measurement. These links included the conduct of research studies that involved: (a) large numbers of students, teachers and schools, (b) methods and procedures for the examination and the analysis of data, (c) the measurement of change over time and across countries, (d) the testing of models and hypotheses for statistical and practical significance, and (e) the drawing of implications from the findings for policy and practice in the field of education.

IEA was extremely fortunate in the quality of the scholars with expert-

ise in educational research that it attracted to work on its programs of research during the initial years, as well as the scholars who came to provide help when their advice was sought. The names of Ben Bloom (United States), Jack Carroll (United States), Torsten Husén (Sweden), Bram Oppenheim (England), Gilbert Peaker (England), Bob Thorndike (United States), and David Walker (Scotland), who collaborated as scholars in the writing of major works, were widely known throughout the world for their contributions to research methodology and measurement. Likewise, Arnold Anderson (United States), Bruce Choppin (England), Georg Rasch (Denmark), Leslie Kish (United States), Douglas Pidgeon (England), John Tukey (United States), Ralph Tyler (United States) and Herman Wold (Sweden), each in his own way, provided help when needed during the early years of IEA's program of research.

At the forefront of activities in the early years were Torsten Husén (Chairman) who had great skills in eliciting support, Neville Postlethwaite (Director), who had remarkable drive and organizational skills, and Richard Wolf (United States) who undertook the pioneering data analysis for the Mathematics study (1964). Gilbert Peaker (England) also had immense experience in the planning and conduct of large-scale survey studies in England, and in the analysis of data within the traditions established by R.A. Fisher. He had the flexibility of thought and the time to devote to solving the many problems that arose in the analyses of the data of both the Mathematics (1964) and the Six Subject (1966-1975) Studies.

The increasing power of technology, particularly of computers and the optical scanning of documents over the 30-year period, made it possible for IEA to collect, store, access and analyse increasingly large bodies of data from classrooms and schools across the world. Moreover, from different parts of the world different approaches have emerged with regard to the statistical analysis of the immense amounts of data associated with hundreds of thousands of cases and hundreds of variables. Not only did the large bodies of data and information attract

scholars who sought to test relationships that helped to explain the different ways education was provided in schools in different countries and regions of the world, but they also led to the development of new procedures for the examination and analysis of the data.

This chapter tells the story of IEA's relationship not only with the methods of research but also with the statistical analysis of data through bringing scholars together from different parts of the world to work both as individuals and in teams to collect, store and examine very large and rich bodies of data. The different traditions of research and statistical analysis have advanced-- in different ways and to different extents-- the efforts made towards the solving of a wide range of problems. This has led not only to new knowledge but also to new approaches to research and new procedures of measurement. The further advantage of the cooperative efforts of so many people from so many different countries was that the new approaches and methods were widely disseminated more rapidly than would otherwise have been possible. Each IEA project since the early 1970s has conducted training sessions on the methods of research and the analysis of data. Moreover, studies have encouraged participants to write and publish their reports in their own countries using the newly developed approaches and analytical procedures.

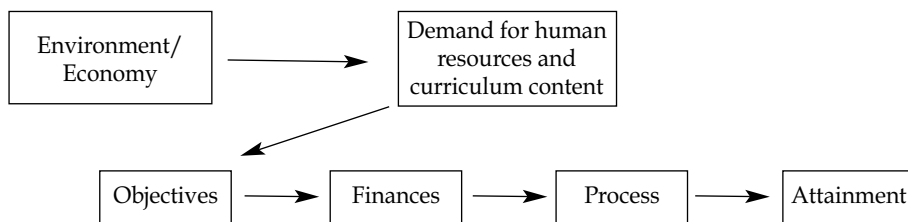
### **The Advancement of Theory and Models of School Learning**

The scholars who met first in Hamburg in 1958 and 1959, and subsequently in England in 1959, to plan a program of cross-national research into the processes of education across different countries and into the provision of educational services in schools that influenced educational achievement, sought initially to build a body of knowledge. The origins of this work are described and discussed in a chapter in the first volume of the two books that mark the 50 years since the founding of the International Association for the Evaluation of Educational Achievement (IEA). The first steps in this work involved

the conduct of a Pilot Project in 1960, to establish that it was feasible to undertake a study across different countries (Foshay et al., 1962). This was followed by a study of achievement in Mathematics in 1964. The design of the Mathematics Study (Husén, 1967) “involved the advancing and testing of specific hypotheses in four main areas: (a) the effects of school organization on mathematics achievement, (b) curriculum and instruction, (c) support for education, and (d) the relationships of social, economic and geographic factors with mathematical achievement” (p. 110).

After the reports of this study were published in 1967, a conference was held at Lake Mohonk in the United States to develop the theoretical foundations for a major study in six subject areas that was already being planned across the three levels of schooling, the 10- year-old level, the 14-year-old level, and the terminal secondary school level. The title of the report of this conference, ‘Towards a Cross-National Model of Educational Achievement in a National Economy’ (Super, 1967b), indicated the nature and the scope of the discussion and purpose not only of the conference, but also of the theoretical foundations that were seen to be required for the next stage in the research program. The variation between countries in their economic circumstances was seen to be related to the differences in the provision of education in each country and to the nature and levels of achievement attained at successive stages of schooling. Urban Dahllöf from the Uppsala University in Sweden in an opening paper (Dahllöf, 1967) developed a simple scheme for the educational process that applied in cross-national settings (see Figure 1).

The ideas that emerged from this conference were complex but coherent, and led to a model (see Figure 2) that involved a flow from Inputs to Processes to Outputs and Utilization set in a framework of Social and Cultural Capital on the left and National and Human Development on the right hand side. Several features of the model are of interest. The financial flow alongside the pupil flow and the analytical or causal flow from left to right in the figure links to the use of



**Figure 1.** A general model for cross-national comparisons in education  
(Source: Dahllöf, 1967)

regression analysis to estimate the magnitudes of the effects. The use of boxes and circles to indicate observed and latent variables, respectively, leads on to the use of path analysis with latent variables. While the levels of operation at the systemic, institutional and individual levels are not ignored, they are not clearly conceptualized. Furthermore, the emphases on 'process and utilization' add components that go beyond simple 'input' and 'output' relationships. The challenge to the research workers involved in the IEA studies was to test this theoretical model and to estimate the magnitudes of the hypothesized effects. Subsequently, a stronger theoretical framework could be developed that would be based on empirical evidence that had been obtained from cross-national research studies. This initial model guided the substantive reports arising from the Six Subject Study conducted by IEA in the years that followed.

Two other models were employed in the building of an understanding of the processes of education that were developed independently. They were used in the reporting of the Mathematics Study conducted in 1964 and the Six Subject Study in 1970-71 and were incorporated within the Lake Mohonk model, namely: (a) model of selectivity and retentivity, and (b) model of school learning.

### *The Model of Selectivity and Retentivity*

Walker (1967) constructed a model for the effects of selection in the comparison of standards of achievement across countries and school

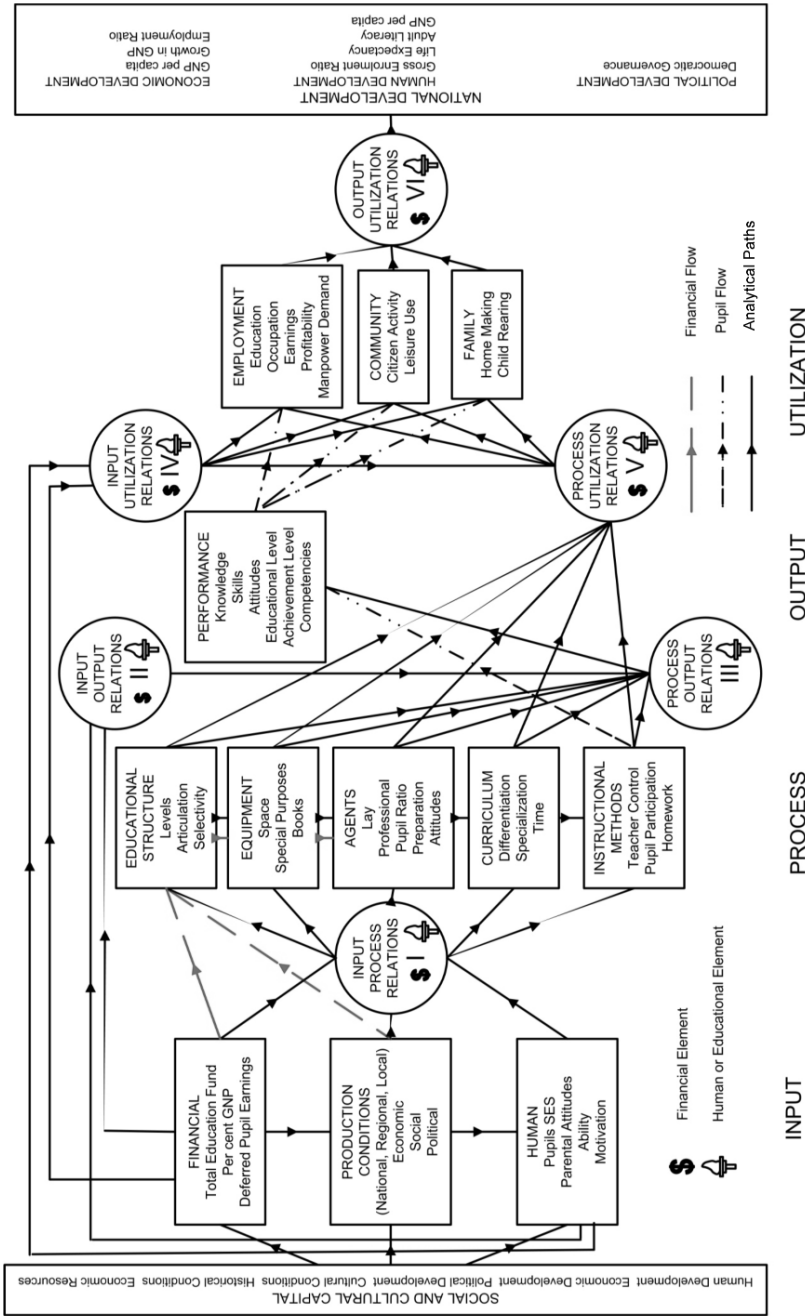


Figure 2. A comprehensive model of educational achievement (Source: Super, 1967a, p. xx(12) modified)

systems where different proportions of the age cohorts were involved. The basic principle underlying this model was that each complete age group in each country had the same distribution of mathematical ability and that the differences in means and variances found at a specified age level, such as the terminal secondary school level were the result of selection or retention procedures. Thus, it could be hypothesized that the higher the retention rate, or the lower the degree of selectivity, the lower would be the average level of achievement in mathematics of the students tested. There were several underlying assumptions involved in the use of this model. The simplest assumptions were: (a) mathematical ability was normally distributed among students in an age cohort within a system and across systems; (b) largely common curricula and methods of instruction operated across systems; (c) those students in the population under survey were the best and had the highest mathematical ability within the system; and (d) the model could be modified by assuming a correlation between the variable operating to select students and the mathematical ability of the students. Through the use of the truncated normal distribution under the assumptions for a specified proportion of the cohort under survey, the mean scores and the variances of the scores were calculated from the model and the model was tested for its adequacy against the data collected in the study. If the model estimates did not fit the observed data, the model would need to be rejected, otherwise it could be accepted, but pending further investigation. Walker (1967) tested such a model using data from the Mathematics Study and support for the model was found.

### *Carroll's Model of School Learning*

The model of school learning advanced by Carroll (1963) was published after the Mathematics Study had been planned. While some relationships associated with time as a variable were reported by Husén (1967), it was only in the Study of French as a Foreign Language that Carroll (1975) was able to test across countries his model of school



learning in 1975. Carroll developed the model in order to examine success on complex learning tasks. The variables were specified in terms of time: (a) **aptitude**, which involved the amount of time a student would require to learn a task to a specified criterion, given motivation, opportunity to learn and optimal quality of instruction; (b) **perseverance**, which involved the amount of time that a student was willing to engage in active learning; and (c) **opportunity to learn**, which involved the amount of time provided for learning in a specific program. In addition, two other variables were not specified in terms of time: (d) **ability to understand instruction**, which was dependent on the (e) **quality of instruction** provided. **Quality of instruction** involved the structuring of the learning task, the effectiveness of presentation and the skills of the instructor. These two additional variables interacted with each other and with the three variables directly involving time. Bloom (1974) drew attention to the relevance of IEA's findings on time for the study of school learning and gradually the significance of time as a key concept in learning came to be widely accepted.

In 1981, during the planning of the Second IEA Science Study (Keeves, 1992b) a causal model derived from Carroll's model of school learning and the Lake Mohonk model was advanced, and was subsequently tested in the analyses of the data collected in that study at the 10-year-old, 14-year-old and terminal secondary school levels. This model of performance in science is shown in Figure 3 and its dependence on Carroll's ideas is apparent. In advancing this model for investigation it was found to be necessary to indicate that some variables were considered to operate at the student level, some variables at the classroom or teacher level, and some variables at the school level. While separate analyses could be carried out at each of the three levels, with substantial problems of aggregation bias and mis-estimated precision in the standard errors of effects, no way was known at the time of planning the study of how to undertake the necessary estimation of the parameters of the model and the testing of alternative models.

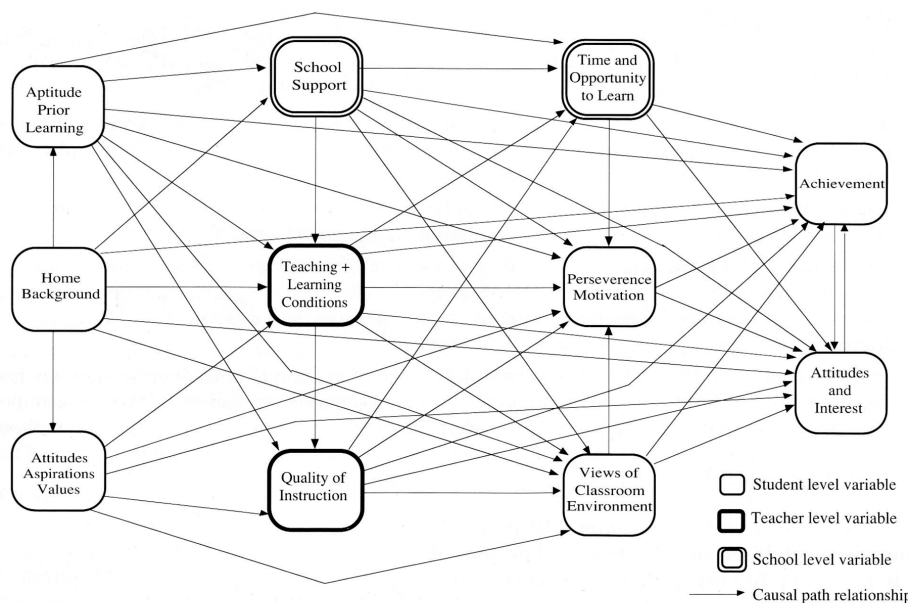
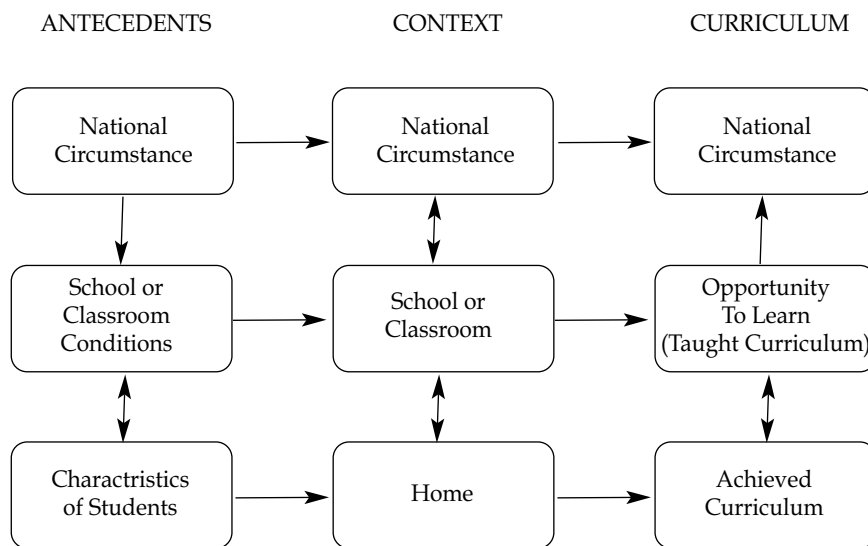


Figure 3. Model of science learning (Keeves, 1992b, p. 45)

### *A Model of Curriculum Implementation*

At the Gränna Workshop conducted by IEA in Sweden in 1971 during the administration of the Six Subject Study, attention was drawn to the problems related to curriculum design and development. A model of curriculum implementation was advanced and tested, at least in part, in the reporting of the results of the First IEA Science Study (Keeves, 1974), the Second IEA Science Study (Keeves, 1992a; Rosier & Keeves, 1991), and the Second International Mathematics Study (Robitaille & Garden, 1989). Figure 4 shows a late version of the model derived from the sources given above.

The curriculum could be considered to operate at three levels: (a) the **prescribed curriculum**, (b) the **opportunity to learn** (taught curriculum), and (c) the **achieved curriculum**, which would be influenced by the antecedent and contextual factors operating at the systemic, school, classroom and student levels, respectively. The **prescribed cur-**



**Figure 4.** The context and components of the school curriculum (Source: Rosier and Keeves, 1991, p.7, modified)

**riculum** was usually specified by authorities responsible in the education system. The **taught curriculum** was located at the second level in the curriculum sequence. It was the task of the individual teacher to interpret the prescribed curriculum by translating it into a set of specific learning experiences that were considered appropriate for the particular group of students in a classroom forming the **opportunity to learn**. The **achieved curriculum** was at the third stage. It referred to the extent to which individual students had learnt from the experiences that were planned and organized for them. Figure 4 shows that the prescribed curriculum was set in the context of the education system, the taught curriculum was located in the context of the school or classroom, and the achieved curriculum occurred in the context of the individual student. Moreover, it was clear that the taught curriculum was influenced by the prescribed curriculum, and the achieved curriculum by the opportunity to learn implemented in the classroom. At the time when this model was first examined some aspects could be considered and effects estimated, but the multilevel structure of the

model could not be examined effectively. Although the relevant data were available, the model had to be considered separately at the student, school and systemic levels.

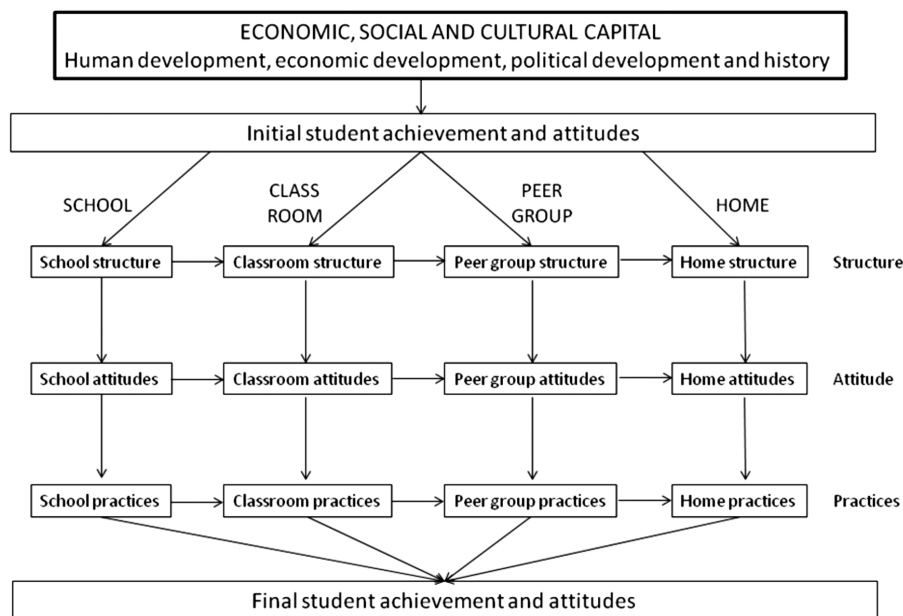
### *A Model of the Educational Environment*

From the discussions at the Lake Mohonk Conference four statements of specific hypotheses emerged, as well as a causal model showing the influences of the educational and social systems, the school, the classroom, the peer group and the home on learning and development at the individual student level. The characteristics of these four facets of the educational environment were argued to form blocks of a structural, attitudinal and behavioural (involving practices) nature that contributed to the outcomes of student achievement and attitudes. A cross-sectional model was clearly inadequate and a longitudinal model was necessary if causal relationships were to be considered. This causal model (Figure 5) was not issued in the published report of the Lake Mohonk Conference but published independently (Keeves, 1972).

## **The Development of Models**

The causal models included not only the direct effects of the different cells on the outcomes of student achievement and attitudes, but also the operation of mediated effects. Gilbert Peaker (1967) in the Plowden Report had established the importance of attitudes over and above the effects of socioeconomic status, while Coleman (1996), in the *Equality of Educational Opportunity Report*, had shown the importance of the socioeconomic status of the school on educational outcomes. Moreover, some of the students in the MESA Program at the University of Chicago under the influence of Bloom (1964), including Dick Wolf, had investigated the effects of the practices of the home on educational outcomes.

From these models and the underlying theoretical perspectives, fur-



**Figure 5.** A model of the educational environment (Keeves, 1972, p.39, modified)

ther models were derived for examination, debate, estimation and testing in the subsequent research programs conducted by IEA. However, the methods of investigation were limited by the analytical traditions and procedures available at the times the studies were conducted. This situation provided a powerful stimulus to find solutions to these problems. The areas in which solutions were sought are considered in the sections that follow.

- (1) There were issues concerned with the framing of specific hypotheses and the making of a decision to endorse or reject the hypothesis or, alternatively, to develop a more complex model that described and explained events occurring in the real world, leading to an examination of the model and alternative models.
- (2) There were issues concerned with the collection of data from a sample that had a complex design, as well as the estimation of the associated parameters with their appropriate standard errors.

- (3) From the use of many observed variables that were interrelated in order to examine their influence on an identified outcome, the need arose to form latent variables that combined the effects of the observed variables, as well as to examine the indirect effects of the latent variables that operated through other latent variables.
- (4) Issues arose in the examination of data from the situation that was present within most analyses in educational research studies, since the data had a multilevel structure.
- (5) In educational research, there was primary concern for learning and development that involved the investigation of change over time. The examination of data in this context not only required the use of time series, longitudinal or trend data but also required that the data were measured on an interval scale across an extended scale that was not truncated at either the higher or lower ends.
- (6) The measurement of achievement, attitudes and background variables needed to produce data of high quality, which required the development of new techniques for scaling and scoring.

### *Hypothesis Testing for Decision-Making or Modelling for Explanation*

For 25 years, Gilbert Peaker guided the sampling designs and the statistical analysis of data collected in the IEA studies. He was initiated into mathematical and statistical analysis at Cambridge University in England and was greatly influenced by the seminal work of Fisher (1925). Fisher advanced what Lindsey (1974), who worked on the secondary analysis of IEA data, subsequently referred to as a 'modelling approach to statistics' (Lindsey, 1995). This approach was consistent with Fisher's approach to scientific inquiry in which a model was rejected if it was found to have less than one chance in 20 of being consistent with data obtained from the real world. Peaker also made similar judgements that were based on the estimated size of an effect that was of practical significance. The alternative approach to statistical

analysis involved the making of a decision between clearly specified alternative hypotheses and it was desirable to reduce the risk in the long term of making mistakes. This approach is referred to as the 'Neyman-Pearson' or 'decision-making' approach and is widely used in North America. Failure to see these two approaches as being concerned with different types of research questions has led to widespread controversy and strong support for a particular approach in different countries and regions of the world.

The decision-making approach has led to the view that in order to argue for a causal relationship a tightly designed experimental study must be conducted. Thus, the procedures of analysis of variance or the analysis of covariance can be employed to investigate learning and development in educational research. This problem was addressed in the IEA research program in the Mathematics Study in 1965 and it was resolved that since it was not possible to undertake experimental or intervention studies across countries, the use of analyses of covariance procedures would generally need to be avoided. Consequently, the modelling approach was adopted and regression analysis procedures were employed. In the regression analysis, statistical control was used to allow for the influence of extraneous factors in the examination of causal relationships.

In a modelling approach causal relationships that involved either a temporal sequence of events, or an argued causal influence, could be built or 'structured' into the model and allowed for in the regression analysis. The shift to a modelling approach also permitted the use of maximum likelihood estimation procedures to be employed in the examination of a model, and the estimation of causal effects. The model could then be rejected if the model were found not to be consistent with the observed or recorded data obtained from the real world. The modelling approach adopted by IEA has been accepted more widely in Europe than in North America and has been promulgated by IEA in those North American centres where IEA or European trained scholars were working.

### *Sample Design and Estimation of Errors*

Gilbert Peaker greatly influenced the design of samples for all IEA studies following the Pilot Project. He emphasized that students were clustered within classrooms, classrooms were nested within schools and schools nested within districts and regions. Consequently, allowance needed to be made for the effects of design and the effective sample size in the estimation of all statistics. Random selection at successive stages was necessary for the estimation of sampling errors. Stratification was also necessary to examine the effects of school type, and in the subsequent analyses, the effects of stratification needed to be examined. Moreover, for every statistic estimated, and for each of the variables involved, these effects of sample design differed. At the time, none of these issues were addressed in courses on educational research, in text books or computer programs. Only in 1965, after testing for the First Mathematics Study had been completed did Kish (1965) publish a book that dealt with these issues. IEA studies employed successively: (a) four independent replications in the Mathematics Study in 1964, (b) jackknifing with ten jackknife units in the Six Subject Study in 1970-71, and (c) jackknifing with the school employed as the jackknife unit in the next round of studies in the 1980s. The last named procedure (c) was based on a study undertaken by Rust (1984) at the University of Michigan. The estimation of standard errors was a major task since it had to be carried out for all estimated parameters as well as national mean scores on the major achievement tests of each study.

The IEA studies exposed the problems associated with the design of cluster samples in educational research and the estimation of the errors of sampling and the testing of findings for statistical significance to a world audience who were largely unaware of the nature and complexity of the problem.

### *Many Observed Variables and the Mediation of Effects*

From the outset in the Mathematics Study, it was recognized that in



each country many variables were operating to influence the teaching and learning of mathematics. Not only was it of interest to identify those variables that were effective, but it was also of interest to identify those variables that were commonly thought to be effective and for which recognizable effects could not be identified. Moreover, since control was not exercised in a quasi-experimental way or through an intervention, it was argued that the use of analysis of covariance was, in general, an inappropriate statistical procedure to use in the examination of the data. In an initial scanning of the data, strip-correlation procedures were used, by correlating the explanatory variables with a small set of selected outcome variables, as well as by partialling out (or stripping) through the use of regression procedures the effects of selected variables, such as a stratifier, verbal ability or prior performance. This enabled the partial regression coefficients to be examined, and those variables with stronger effects could be included in further analyses.

In the Mathematics Study, the stepwise regression analysis procedure was used that had been advanced by Thomson (1951) of the University of Edinburgh, with the analysis undertaken at the student level, and with the classroom, teacher, school and system variables disaggregated to their lowest level. Cut-points for statistical significance were estimated from the replicated analysis of the four sub-samples provided by each of the countries in the packaging of their data for forwarding to the data processing centre in Chicago, where Dick Wolf and his colleagues undertook the work of analysing the data.

In the Six Subject Study, it was recognized that grouping together significant variables into blocks was necessary in order to avoid suppressor effects, first reported by Thorndike (1949). This procedure was subsequently termed 'block-wise' regression analysis (Pedhazur, 1982). The use of canonical analysis to form latent variables was suggested and in reporting the Study of French as a Foreign Language, Carroll (1975) employed this procedure. It was also suggested that the mediation of effects in a causal sequence could be provided for by using path

analysis. The use of path analysis in educational research was first employed in 1971 by Peaker (1971) in a report on "The Plowden Children Four Years Later". A simple example of path analysis was reported for the First IEA Science Study (Comber & Keeves, 1973) undertaken in 1970-71.

The statistical analyses for the Six Subject Study (Peaker, 1974) were carried out in Stockholm under Gilbert Peaker's and Neville Postlethwaite's supervision, with other scholars coming to Stockholm to provide help and advice. Among the Swedish scholars who became interested in the issues were Jöreskog and Wold. They saw that the analytical problems which arose in the IEA studies provided situations where the ideas of path analysis that involved latent variables could be employed. Jöreskog had developed his ideas of confirmatory factor analysis and structural equation modelling using maximum likelihood estimation procedures, and Wold his idea of partial least squares regression analysis with latent variables that involved the principles of canonical analysis (Jöreskog & Wold, 1982). Munck (1979) applied Jöreskog's Linear Structural Relations (LISREL) model, and Richard Noonan (1976) used Wold's Path Analysis with Latent Variables (PLS) model in the analyses of IEA data in the Stockholm Institute adjacent to the IEA Headquarters in the University of Stockholm. These two secondary data analysis studies were examples of the completely new field of multivariate statistical analysis with a wide range of applications (see, e.g., Jöreskog & Wold, 1982). Each approach developed its followers, with LISREL having adherents mainly in psychology and PLS having its disciples in economics, business and commerce. Thus, educational research workers from around the world were learning to use both approaches since they essentially had come to serve different purposes. LISREL was primarily confirmatory, with latent variables used in the reflective mode. PLS was largely used in an exploratory way, with latent variables being formed either in a formative mode or a reflective mode.

New ideas about the analysis of data were spread across the world by the Spencer Fellows who worked on the secondary analyses of IEA

data at the Institute headed by Torsten Husén at the University of Stockholm. Among the Spencer Fellows was Jeremy Finn from the MESA Program at the University of Chicago, who also developed an analysis of covariance program titled 'Multivariate' (Finn, 1972) that, as its name suggests, could be used with large multivariate sets of data. Jeremy Finn and Ingrid Mattsson (1974) illustrated the use of this program in the examination of IEA data.

### *Multilevel Analysis and the Moderation of Effects*

While IEA scholars were generally aware of the consequences of the clustered nature of the data involved in educational research and, in particular, in survey research, they were less aware of the effects arising from aggregation bias in the estimation of correlation coefficients. Gilbert Peaker had provided for the effects of sample design in significance testing and in the estimates of standard errors. While Bob Thorndike's father (E.L. Thorndike) had warned against aggregation effects in 1917, consideration of the problems associated with the consequences of the aggregation and disaggregation of data relating to the estimation of school and classroom treatment effects had not received any consideration outside of analysis of variance with a very rigidly balanced sample design. In the Mathematics Study, the data were analyzed at the student level only. However, in the Six Subject Study of 1970-71, Gilbert Peaker from his experience in the analysis of the Plowden Study data (Peaker, 1967) argued for the examination of the data separately at the school and student levels with regression analysis. The strategy employed aggregated or disaggregated data, respectively, with some allowance made for the effects of sample design through the use of the jackknife procedures for the estimation of standard errors in significance testing. Only after the publication of a seminal paper on the use of weighted least squares in late 1971 were the school level analyses recomputed with each school sample weighted by the size of the sample, in addition to the weighting of the data for the sample design. This required that in the final report of the First IEA

Science Study, the presentation and discussion of the results of this school level regression analysis (Comber & Keeves, 1973) had to be rewritten just prior to publication. Nevertheless, the more fundamental issues associated with the levels of operation at the student and school levels were discussed informally at considerable length and appropriate procedures that avoided aggregation bias effects could not be envisaged, although partitioning of variance at the student and school levels were clearly required.

Prior to the next round of studies that included the Second IEA Science Study and the Classroom Environment Study, the work undertaken at the University of Michigan by Mason, Wong and Entwisle (1983) became known. This study employed maximum likelihood estimation procedures and resolved key aspects of the problem. This procedure was replicated with least squares regression analysis using data collected several years earlier in relation to the First IEA Science Study. In an examination of the effects of class size, meaningful and coherent results (Larkin & Keeves, 1984) were obtained. The approach employed considered at the macro-level the slopes and intercepts formed at the micro-level as outcomes, with macro-level variables tested for their moderating influence on micro-level effects. This added a completely new dimension to regression analysis, although it was recognized that maximum likelihood estimation would be more effective than least squares regression. With the assistance of Ray Adams, this work paved the way for the resolution of a major problem that also arose in the analysis of data for IEA studies which involved the estimation of moderating or cross-level interaction effects.

The significance of the work undertaken at the University of Michigan had been recognized in the United States and the Spencer Foundation funded developmental work that was undertaken by Tony Bryk and Richard Congdon in the Department of Education at the University of Chicago in collaboration with Raudenbush at the University of Michigan. Subsequently, work was carried out in the IEA centre at the University of Hamburg to examine data collected in Hong Kong as

part of the Second International Mathematics Study by Cheung, Keeves, Sellin, and Tsoi (1990). This study, reported by Cheung and his colleagues compared the results obtained with the Hierarchical Linear Modeling program (HLM) developed by Bryk and Raudenbush (Bryk, Raudenbush et al., 1988) with an alternative program that had recently become available developed by Goldstein (1987) at the University of London, which employed generalized least squares with a general linear model strategy. While the HLM procedure was preferred as a possible approach for the analysis of the Second IEA Science data the computer facilities and staff available to carry out the extensive work required were inadequate. Nevertheless, a major problem had been resolved for the examination and analysis of IEA data sets in the future.

### *Investigating Change over Time in IEA Studies*

Change over time has been examined in several ways in IEA studies. There were generally considered to be five different design strategies for longitudinal studies, namely simultaneous cross-sectional studies, trend studies, time series studies, panel studies and intervention studies. The IEA studies undertook, in the main, simultaneous cross-sectional studies in which different age or grade groups were examined with different samples at the same time point, for example the Mathematics Study in 1964. However, the two Science Studies conducted in 1970 and 1984 were purposefully designed as trend studies in which the same age groups were examined with different samples at two different time points. The study conducted by Keeves (1972) in 1968 and 1969 in the field of Science in conjunction with the First IEA Science Study was a time series study with the same sample examined on two occasions at successive grade levels that was sometimes simply referred to as a 'longitudinal' study. All three types of longitudinal studies compared test performance between occasions or stages of schooling where the achievement tests were generally different to provide for minor changes in curricula or for changes in the levels of

school learning involved. Moreover, if only two points in time were involved, absolute change could generally not be examined because of the low reliabilities of the difference scores. Under these circumstances relative change was examined and regression analysis procedures were employed. If, however, three or more time points were involved absolute change could be examined through the use of hierarchical linear modelling, with both intercepts and slopes taken as outcomes at a higher level than the scores.

Nevertheless, in most approaches to examine change it was inappropriate in educational studies to employ the same test on successive occasions. This was partly because of the possibility of practice effects and partly because it was difficult to use so-called secure tests, but largely because changes in the curricula and learning experiences provided were inevitable.

It was clearly unsatisfactory to use the same tests across age and grade levels because of curricular differences. Under these circumstances, different tests needed to be equated both across levels and across occasions. Regression analysis procedures had generally been employed to provide an appropriate way to equate tests across age or grade levels or alternatively across occasions. However, Rasch scaled scores could be employed that were independent of the items or tasks involved and independent of the persons used to calibrate the tests, provided that there were adequate numbers of link items used. Thus, Rasch scaling could be used not only to measure performance on an interval scale but also to equate tests both across age and grade levels by vertical equating, as well as within age and grade levels by horizontal equating across occasions. This was evident to some of the research workers developing the IEA studies from the outset, as was indicated by the events that followed their discussions with Rasch in Hamburg in 1958 (see later section) and the lectures given in Chicago in 1960.

### *Scaling and Scoring in IEA Research*

Over decades, it has become increasingly apparent that in educational

research, while the normal distribution still had its place in the assessment of human abilities, the Poisson, the binomial and the multinomial distributions were involved in the generation of much of the data collected in IEA studies. Thus, the data analysis was concerned with the study of events that occurred to independent individuals or involved frequencies of independent events. Where the same types of events occurred repeatedly to the same individuals, counts of dependent events were involved. Time to perform a task was a rare case in educational studies where there was an underlying continuous normally distributed activity, for which many of the available statistical procedures had been developed in order to analyse the data collected in educational research studies. These changes in the development of the statistical analysis of data had led to marked changes in the ways in which most research scholars in IEA studies went about their work. This section discusses some of the changes that occurred with respect to the assessment of performance on achievement tests, the quantification of information on attitudes and views, the scaling of data obtained from the classification of information into categories, and the observation of events in the classroom.

### **Assessment of performance involving guessing**

In the Mathematics Study, all assessment of performance on achievement tests merely involved the classification of responses to items and tasks as either 'right' or 'wrong' and the counting of the number of correct responses. In the Six Subject Study in 1970-71, similar scoring procedures were involved for responses to items and tasks that were grouped into three categories of 'right', 'wrong' or 'omit', and as a correction for the possibility of a person guessing in the choice of a 'right' or 'wrong' response category, a 'correction for guessing' was applied. Bruce Choppin (1974) wrote an IEA monograph in 1974 that examined the issues concerned with the use of a correction for guessing in test scoring. Bob Thorndike (1982) also wrote a chapter on this issue in his book titled *Applied Psychometrics*. This correction was introduced to take into

account more appropriately differences between countries with respect to students' tendencies to guess. Moreover, the use of a 'guessing correction' was argued to provide scores with greater variance and scores that were better suited to the multivariate regression analysis procedures that were subsequently employed in the analyses of the data.

### **Performance tasks of speaking, writing essays and practical work**

In the Mathematics Study in 1964 constructed response items were administered but were simply scored 'right' or 'wrong' with an entry made on the Optical Mark Sensed Answer sheets that were subsequently machine scored together with the multiple choice items. In the Six Subject Study in 1970-71, the Science, Literature, French and English Studies included performance components that were not machine scored and subjective judgements had to be made, commonly on a scale ranging from 0 to 4. However, Carroll (1975) reported from the French as a Foreign Language Study:

In the present study, it can be said on the basis of the data now analysed that the correlations between scores on "objective type" tests and scores in free-response, subjectively-scored tests are often so high that it is difficult to believe that testing procedure ("objective" vs "subjective") makes any practical difference in the results. (pp. 52-53)

Subjective type items were normally scored by one person, but in order to estimate the inter-rater reliability, some items were independently scored by two or more people. Bob Thorndike (1982) subsequently addressed the issue of estimation of reliabilities of multiple tasks and multiple raters and the interactions between persons, tasks and raters from the perspective of generalizability theory in a chapter in his book on Applied Psychometrics prior to the next round of IEA studies in the 1980s.

### **Criterion scaling of categorical variables**

A major task in the analyses of data for the Six Subject Study in the



early 1970s involved the statistical control for the effects of the stratification of the samples employed. While stratification was used to obtain better estimates of mean values, and needed to be taken into consideration both in the weighting of the data and the estimation of sampling errors, it also operated to modify the variability of the scores recorded. Consequently, in any subsequent analysis of data using multivariate analysis procedures, the stratifying variables needed to be included in the analyses and control exercised for their effects. In addition, it was commonly of interest to examine the differences in performance by students and schools in the different cells of the strata. Thus, the strata formed categories and an appropriate variable needed to be formed out of the several stratifying categories. Gilbert Peaker required that this should be done for all samples in all analyses through the use of a procedure that became known as 'criterion scaling' that had been developed by Beaton (1969). Since the cases in each of the categories of the strata were treated independently, a new variable could be formed by weighting each of the categories using either: (a) mean score weights or (b) regression score weights.

This procedure was also extended to form ordinal variables out of categorical variables using what were termed 'rank scaled scores'. Appropriate rank scaled values were generally integral in nature and centred around a zero that operated as 'dummy' in the regression estimation of such scores. These scaling procedures were widely employed in the Six Subject Study analyses. In the Mathematics Study in 1964 the variable of socioeconomic status was assigned rank scaled scores of 1 to 9, but with some misgivings: "One might therefore have doubts about the meaningfulness of calculating the means and standard deviations of occupational distributions" (Husén, 1967, vol 2, p. 110). Nevertheless, it became standard practice within a few years to use criterion scaled and rank scaled scores. Thus, father's occupation operated as a variable with regression and correlation coefficients readily calculated using scaled scores from 1 to 9 that were validated by comparisons of the rank score values with performance levels in

mathematics achievement at the 13-year-old age level (Husén, 1967) in spite of the misgivings expressed above.

### **Socio-educational advantage and disadvantage**

As discussed in the previous section, the Mathematics Study introduced issues about the effects of the socioeconomic status of the occupation and the level of education attained by the parents of the students under survey using rank scaled scores. The Six Subject Study employed similar questions concerned with level of education and level of occupation but permitted each country to employ its own classification of levels that were subsequently criterion scaled. In addition, simple questions about the homes of the students were asked, such as the number of books in the home, family size and availability of daily newspapers. In the second round of studies in the 1980s questions about the possessions relating to the wealth of the home were sometimes asked. The scores assigned at the individual level were aggregated to the classroom, school and systemic levels to provide an index of climate of the communities from which students were drawn. The findings from such information obtained from multivariate analyses at the student and school levels attracted considerable attention, but not without controversy in some countries where issues involved in equity of educational opportunities were of concern. Thus, issues associated with socioeconomic status led to some consideration of the economic level of the countries under survey in IEA studies, although none of the participating countries could be considered to be developing countries in the First Mathematics Study. Interest in economic issues arose in the IEA studies undertaken after the First Mathematics Study that included some developing countries. The findings of these studies reported the marked divide between the two groups of developing and developed countries. However, little attempt was made to identify the socio-cultural and socio-educational factors that gave rise to this divide, implying that the factors involved were essentially economic in nature.

This led Ingemar Fagerlind, who followed Torsten Husén as Director of the International Institute of Education in Stockholm, together with Lawrence (Larry) Saha, a visiting fellow at the Institute, to prepare a volume titled, *Education and National Development: A Comparative Perspective*, that was concerned with education in developing countries (Fagerlind & Saha, 1989).

### **Rasch scaling**

At the initial meeting of the founders of IEA in Hamburg in 1958, Georg Rasch from Copenhagen was present, but he did not attend subsequent meetings for the planning of the Pilot Project or the Mathematics Study. In the report of the meeting in March 1958 by Fernand Hotyat the following statement was made:

The attempt has also been made to grade tests irrespective of the age of the pupils. Thus Dr Rasch of Denmark has designed attainment tests which are graded according to the intrinsic progressive complexity of subject matter; the marks, in this case, aim to assess the absolute level which the child has reached in a particular subject. (quoted in Postlethwaite, 2009, p.4)

The importance of Rasch's work would have been immediately evident to David Walker from Edinburgh on whose work his colleague Lawley (1943) had drawn to advance the idea of conjoint measurement. Lawley had also proposed that achievement test performance could be assessed on a scale of measurement based on the normal distribution, and had undertaken the detailed computation by hand to illustrate how such a scale could be formed. At that time, the possibility of measurement in the behavioural and social sciences was rejected because the performance on a test item or task was argued to be dependent on the relative underlying difficulty of the tasks. Lawley indicated that if measurement was based on a term  $(x-a)$ , where 'x' represented the ability of a person and 'a' the difficulty of an item on a latent trait, and the difference was assumed to be normally distributed, a scale of measurement could be constructed as a conjoint scale.

Moreover, if Thurstone's idea that responses by the students were stochastic in nature was accepted, then probabilities were necessarily involved in the process of measurement. Unfortunately, the estimation of scores using the normal distribution involved extremely heavy computation and this strategy for measurement in the fields of education and psychology awaited advances in computation. Lawley (1940) had already shown that the necessary estimation in factor analytical work could be readily carried out using maximum likelihood procedures instead of least squares procedures, and this helped solve the problems of computation.

During the 1950s, Rasch undertook work on the scaling of examination performance in the Danish school system that used the logistic function instead of the normal distributional function employed by Lawley in the construction of a scale of measurement. The important characteristic of a measurement scale in education was that an underlying interval scale should be formed so that changes over time across different levels of the scale were equivalent and could be analysed. In addition, the measures estimated should be independent of the items and the persons employed to calibrate the scale. This involved the condition that the items and persons were consistent with the requirements of the scale as indicated by the item characteristic curve of the logistic function. This approach required that the logarithm of the odds of a person's ability relative to the difficulty of an item was associated with a correct response.

Georg Rasch was invited in 1960 to lecture in the Measurement, Evaluation and Statistical Analysis Program in the Department of Education at the University of Chicago, and a book on his work was published in 1960 by the University of Chicago Press (Rasch, 1960). Subsequently, Lord from the Educational Testing Service (ETS) in Princeton in the United States developed an alternative approach that incorporated allowance for items that did not fit the requirements of the simple one parameter logistic function. In addition, Lord made allowance for items that appeared to involve guessing by some stu-

dents, on the assumption that guessing was a characteristic of an item and not the propensity of a student, as was provided for in the correction for guessing procedure discussed above.

In the Mathematics Study in 1964 (Husén, 1967), the equating of the tests across the three levels of schooling that were the target populations in the study, namely the 13-year-old level, the 15-year-old level and the terminal secondary school level, was planned to be accomplished by the use of several mathematics tests at each level. Tests A(1), B(2), C(3) were administered at the 13-year-old level; Tests 3, 4, and 5 were given at the 15-year-old level; Tests 5, 6, and 7 were administered to the non-mathematics students at the terminal secondary school level; and Tests 5, 7, 8 and 9 were given to the mathematics specialist students at the terminal secondary school level. Thus, between the different levels there was at least one common test that could be used for equating purposes. However, the equating of the tests was neither used nor discussed in the major reports of the study. Nevertheless, one national report (Keeves, 1968) examined the growth between the Grade 8 level and the mathematics specialist students at the terminal secondary school level and showed that for the subgroups based on States the growth in achievement was clearly related to retentivity.

For the Six Subject Study in 1970-71 it was planned that Bruce Choppin, who had worked on the development of Rasch scaling in the MESA program at the University of Chicago, would take charge of a data processing unit under the guidance of Bob Thorndike and Dick Wolf at Teachers College, Columbia University in New York. The achievement tests and the attitude scales would be Rasch scaled and equating would be undertaken across the 10-year-old, 14-year-old and terminal secondary school levels. However, the pressure of work was so great that this plan was abandoned and guessing-corrected scores were employed instead. Nevertheless, in the next round of IEA studies in the 1980s, the Second IEA Science Study planned to undertake the Rasch scaling of the data. Again, the task of cleaning the data files proved to be very heavy and work was delayed because of lack of

funding. Subsequently, Andreas Schleicher undertook the work in Hamburg for all the Science tests, but for only 10 countries, using programs obtained from Chicago. However, anomalous effects were found in the vertical equating between the 14-year-old and terminal secondary school levels and it was argued that the program was not doing what it was purported to do in the handling of missing responses to items. Consequently, the program had to be redeveloped in Chicago and much of the Rasch scaling analysis reworked to carry out equating across levels. The published results were meaningful and the power of Rasch scaling was strongly supported.

Under Al Beaton's guidance, work continued with the Rasch scaling of the data on the reading tests in the Reading Literacy Study (1990-91) with considerable success, but without equating the data from the studies conducted in 1970-71 and 1990-91. This was subsequently done by Lietz (1996) and highly meaningful results were obtained.

As an indication of his commitment, Thorndike (1982) not only employed Rasch scaling in the calibration of the Stanford Binet tests that were released in the early 1980s but he also wrote a clear and simple statement on this measurement procedure in his book, *Applied Psychometrics*.

### **Attitude and View Scales**

In the Mathematics Study, the attitude and view scales were examined using Guttman scaling procedures and in the Six Subject Study, Likert scaling procedures were widely used. However, students in the MESA Program at the University of Chicago developed Rasch scaling programs for use with attitude and view scales. A Rating Scale Analysis Program was developed by Andrich (1978) and the Partial Credit Program was developed by Geoff Masters (1982). These programs were not used in the period under review in this chapter for the scaling of attitudes and views in IEA studies. Guttman scaling procedures were used in the Mathematics Study and Likert procedures were used in the Six Subject Studies.

## Conclusion

Five major research centres were formed that were linked together by the IEA scholars and their IEA activities. These centres, in an alphabetical order, were Chicago, Edinburgh, Hamburg, New York and Stockholm. In Chicago, Ben Bloom through the MESA program in the Department of Education at the University of Chicago supported the development of: (a) scaling of Socioeconomic Status, (b) Hierarchical Linear Modeling, (c) Rasch Scaling; and (d) through Dick Wolf, the use of stepwise regression.

In Edinburgh, the Scottish Council for Research in Education together with colleagues in the University of Edinburgh introduced: (a) the use of maximum likelihood estimation procedures in multivariate analysis developed by Lawley (1940), (b) conjoint measurement through Lawley (1943) and David Walker's own work that led to Rasch scaling, (c) the model of retentivity (Walker, 1967) and (d) the analytical procedure for stepwise regression developed by Thomson (1951).

At the University of Hamburg in Germany, Neville Postlethwaite was responsible for (a) introducing rigour into the processing of educational data through the work of Norbert Sellin, Dirk Hastedt and Heiko Sibberns, (b) the further development of Partial Least Squares Path Analysis by Norbert Sellin, (c) the introduction of Rasch Scaling for the equating and scaling of large-scale data sets through the work of Andreas Schleicher (in Keeves, 1992) and Petra Lietz (1996), and (d) the emphases on policy-oriented research and the monitoring of educational achievement.

In New York at Teachers College, Columbia University, Bob Thorndike contributed through his support for (a) regression analysis, (b) correction for guessing in test scoring, (c) Rasch scaling, (d) estimation of reliability through the use of generalizability theory for performance testing; (e) through Harry Passow and Donald Super for the advancement of theoretical models for the conduct of cross-national research; and (f) through Dick Wolf who monitored the scientific standards of

IEA research studies over a very long period.

In Stockholm, Torsten Husén in the Institute of International Education at the University of Stockholm led IEA, and supported the use of (a) Wold's Path Analysis with Latent Variables and (b) Jöreskog's Linear Structural Relations model. In addition, the Institute contributed to the spreading of empirical research across the world through the Spencer Fellowship Program and the preparation and publication of many doctoral theses based on IEA studies. In addition, specific interest was given within the Institute to the problems faced by developing countries.

All these developments were associated with pioneering work that was based on empirical studies and new methods of research and measurement in the field of education. Furthermore, the *International Encyclopedia of Education* in the two editions that were edited by Torsten Husén and Neville Postlethwaite disseminated the many different aspects of educational research across the developed and developing countries of the World. Above all, the work of Gilbert Peaker in England must be recognized for adapting the ideas of RA Fisher (1925) to guide IEA through its formative years with a modelling approach to scientific inquiry in educational research, and the examination of models together with the estimation and testing of the parameters of the models to represent the processes of education in the real world.

## References

- Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika* 43, 561-573.
- Beaton, A. E. (1969). Some mathematical and empirical properties of criterion scaled variables. In G. W. Mayeske (Ed.), *A study of our nation's schools*. Washington, DC: Government Printing Office.
- Bloom, B. S. (1964). *Human characteristics and student learning*. New York: Wiley.



- Bloom, B. S. (1974). *Time and learning*. *American Psychologist* 29, 682-688.
- Bryk, A. S., Raudenbush, S. W., Seltzer, M., & Congdon, R. T. (1988). *An introduction to HLM computer program and user's guide. Version 2*. Chicago: University of Chicago.
- Carroll, J. B. (1963). A model of school learning. *Teachers College Record* 64 (8), 723-733.
- Carroll, J. B. (1975). *The teaching of French as a foreign language in eight countries*. Stockholm: Almqvist and Wiksell.
- Cheung, K.C., Keeves, J. P., Sellin, N., & Tsoi, S. C. (1990). The analysis of multilevel data in educational research. Studies of problems and their solutions. *International Journal of Educational Research* 14, 217-319.
- Choppin, B. H. (1974). *The correction for guessing on objective tests*. Stockholm: IEA.
- Coleman, J. S. et al. (1966). *Equality of educational opportunity*. Washington, DC: National Center for Educational Statistics, U.S. Government Printer.
- Comber, L. C., & Keeves, J. P. (1973). *Science education in nineteen countries*. Stockholm: Almqvist and Wiksell.
- Dahllöf, U. (1967). Relevance and fitness analysis in comparative education. In D. E. Super (Ed.), *Towards a cross-national model of educational achievement in a national economy* (pp. 1-25). New York: Teachers College, Columbia University.
- Fagerlind, I., & Saha, L. J. (1989). *Education and National Development: A Comparative Perspective* (2nd edition). Oxford: Pergamon Press.
- Finn, J. D. (1972). *MULTIVARIANCE: Univariate and multivariate analysis of variance, covariance and regression*. Ann Arbor, MI: National Educational Resources.
- Finn, J. D., & Mattsson, I. (1974). *Multivariate analysis in educational research. Some applications*. Stockholm: Institute for International Education, Stockholm University.
- Fisher, R. A. (1925). *Statistical methods for research workers*. Edinburgh:

- Oliver and Boyd.
- Foshay, A. W., Thorndike, R. L., Hotyat, F., Pidgeon, D. A., & Walker, D. A. (1962). *Educational achievements of thirteen-year-olds in twelve countries*. Hamburg: Unesco Institute of Education.
- Goldstein, H. (1987). *Multilevel models in educational and social research*. New York: Oxford University Press.
- Husén, T. (Ed.). (1967). *International study of achievement in mathematics* (2 vols). Stockholm: Almqvist and Wiksell; and New York: Wiley.
- Jöreskog, K., & Wold, H. (1982). *Systems under indirect observation. Parts I and II*. Amsterdam: North Holland Press.
- Keeves, J. P. (1968). *Variation in mathematics education in Australia*. Melbourne: Australian Council for Educational Research.
- Keeves, J. P. (1972). *Educational environment and student achievement*. Stockholm: Almqvist and Wiksell.
- Keeves, J. P. (1974). On the evaluation of science teaching. The IEA science project, Deutsche UNESCO-Kommission, Köln, *Implementation of curricula in science education* (pp. 158-178). München: Pullach.
- Keeves, J. P. (Ed.). (1992). *The IEA study of science III: Changes in science education and achievement: 1970 to 1984*. Oxford: Pergamon Press.
- Kish, L. D. (1965). *Survey sampling*. New York: John Wiley.
- Larkin, A. I. & Keeves, J. P. (1984). *The class size question. A study at different levels of analysis*. Melbourne: Australian Council for Educational Research.
- Lawley, D. N. (1940). The estimation of factor loadings by the method of maximum likelihood. *Proceedings of the Royal Society of Edinburgh* 60, 64-82.
- Lawley, D. N. (1943). On problems connected with item selection and test construction. *Proceedings of the Royal Society of Edinburgh* 61, 773-287.
- Lietz, P. (1996). *Changes in reading comprehension across cultures and over time*. Münster and New York: Waxmann.

- Lindsey, J. K. (1974). A re-analysis of class-size and achievement in the IEA mathematics study. *Comparative Education Review* 18, 314-326.
- Lindsey, J. K. (1995). *Introductory statistics: A modelling approach*. Oxford: Oxford University Press.
- Mason, W. M., Wong, G. V. & Entwisle, B. (1983). Contextual analysis through the multilevel linear model. In S. Leinhardt (Ed.), *Sociological methodology 1983-84*. San Francisco, CA: Jossey Bass.
- Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika* 47 (2), 149-174.
- Munck, I. M. E. (1979). *Model building in comparative education*. Stockholm: Almqvist and Wiksell.
- Noonan, R. D. (1976). *School resources, social class and student achievement*. Stockholm: Almqvist and Wiksell.
- Peaker, G. F. (1967). The regression analyses of the national survey. In: Central Advisory Council for Education. *Children and their Primary Schools: A Report of the Central Advisory Council for Education*. Vol. 2. *Research and Theories*. London: HMSO.
- Peaker, G. F. (1971). *The Plowden children four years later*. London: National Foundation for Educational Research.
- Peaker, G. F. (1974). *An empirical study of education in twenty-one countries: A technical report*. Stockholm: Almqvist and Wiksell.
- Pedhazur, E. J. (1982). *Multiple regression in behavioural research: Explanation and prediction* (2nd edition). New York: Holt, Rinehart and Winston.
- Postlethwaite, T. N. (2009). Memories of IEA achievement surveys. In B. Matthews & J.A. Gibbons (Eds.). *The process of research in education*, pp. 1-20. Adelaide: Shannon Research Press.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Chicago: University of Chicago Press (reprinted 1980).
- Robitaille, D. F., & Garden, R. A. (1989). *The IEA study of mathematics II. Context and outcomes of school mathematics*. Oxford: Pergamon.

- Rosier, M. J., & Keeves, J. P. (Eds.). (1991). *The IEA study of science I: Science education and curricula in twenty-three countries*. Oxford: Pergamon Press.
- Rust, K. F. (1984). Techniques for estimating variances for sample surveys. Ph.D. thesis, University of Michigan.
- Super, D. E. (Ed.). (1967). *Towards a cross-national model of educational achievement in a national economy*. New York: Teachers College, Columbia University.
- Thomson, G. H. (1951). *The factorial analysis of human ability*. London: University of London Press.
- Thorndike, R. L. (1949). *Personnel selection: Tests and measurement techniques*. New York: Wiley.
- Thorndike, R. L. (1982). *Applied psychometrics*. Boston, MA: Houghton Mifflin.
- Walker, D. A. (1967). Mathematical achievement and school retentivity. In T. Husén, (Ed.), *International study of achievement in mathematics* Vol. 2, (pp. 116-135). New Work: Wiley.



## CHAPTER 11

# **IEA: International Studies, Impact and Transition\***

*Hans Wagemaker*

The year of IEA's 45th anniversary seems an appropriate time to reflect on what IEA, as an organization, has accomplished. Such reflection also seems appropriate at IEA's first research conference, which has as its focus the secondary analysis of data generated by the IEA studies.

My brief, when I was asked to give this keynote address, was to reflect on the impact of the research that has been conducted under the auspices of IEA on education (broadly considered) since its foundation. The question of impact raises, of course, the issue of how one might judge impact, and what evidence can be used and is available to judge it. In effect, the question that needs to be answered is how effective has IEA been as an organization?

As a research organization focused on education it is reasonable to expect that the work of IEA should contribute to our understanding of teaching and learning, to the educational policy debate within countries and internationally, and, ideally, to research practice itself. In this address, I therefore would like to consider these impact-related questions by taking into account some of the broader historical changes and other imperatives that have occurred over the last 45 years, and the influence that these have had on the work of the IEA.

### **Historical Development**

As most of you know, IEA was founded in 1958. Although its first meetings were held under the auspices of UNESCO, its roots were

---

\* Keynote address at the first IEA International Research Conference (IRC-2004).

firmly tied to academic institutions through its first scholars, such as Benjamin Bloom, Arnold Anderson, Robert Thorndike, and Bill Wall. In its early years, the work of the IEA was organized primarily by these leading academics, who contributed their time and resources on a voluntary basis and whose motivation to a large extent was guided by a desire to understand and identify those factors that might have meaningful and consistent influences on educational outcomes. The international focus was also, for them, a context from which the naturally occurring variation among countries in terms of practice and policy would provide a source of insight for policy reform and improvement. They argued that evidence from across a wide range of educational systems would be of sufficient variability to permit the revelation of important relationships that would otherwise escape detection. Foshay, Thorndike, Hoytat, Pidgeon, and Walker (1962), the authors of one of the first publications to arise from the work of IEA, eloquently expressed this focus in making the case for international comparisons: "If custom and law define what is educationally allowable within a nation, the educational systems beyond one's national boundaries suggest what is educationally possible" (p. 2).

While much of what international studies like those carried out by IEA do is to describe "what is" in terms of how education is practiced in a country (the within-country perspective), the power of such studies is most fully realized when the international context they provide is considered (the between-country perspective). Given the differences in the ways in which education is organized and practiced across cultures and societies, a comparative perspective not only enables an understanding of its many forms but also serves to expand a nation's horizon as to what might be possible.

Identifying models or practices of education from countries around the world as a means of reflecting on one's own practice and experience was—and arguably still is—a key function of international comparative studies and the work of IEA. IEA, more so than any other organization, has brought an international perspective to the work of

educational policy analysis and research. The work of Foshay et al. (1962), for example, can be seen as one of the early attempts to address, and perhaps silence, data-free polemics on the relative merits or otherwise of one education system as opposed to another.

While the early interests of those who worked on IEA studies were largely driven by more academic concerns, such as describing models of practice or understanding the relationship(s) among variables that might be related to educational performance, the concerns of policy-makers eventually became a more significant factor in shaping the work of IEA. Writing in the early 1970s, Postlethwaite (1974) noted: "At all levels in an educational system, from the teacher in the classroom, through the administrator to the policymaker, decisions have continually to be made most of the time on the basis of very little factual information" (p.7). He identified what was to be a central concern of policymakers, namely, that educational policy is formulated and implemented at all levels of the education system, even where system-level constraints, such as a centralized curriculum, restrict what schools and teachers might do. Discretion at the school and classroom level always remains. The question as to how and on what basis policymakers, administrators, and teachers make decisions in the educational arena was to become the concern of comparative studies of education in general and the work of IEA in particular.

Until the early to mid 1980s, the practice of large-scale assessment, internationally, remained rather ad hoc and was characterized by uncertainty as to regularity of assessments. It was, however, during this period that most of the developed countries of the world initiated or experienced significant reforms in education and the wider public sector (The World Bank, 1999). Similarly, in many low- to middle-income countries, educational reform as a means of enhancing social and economic wellbeing received increasing attention. It was also during this period that the work of IEA can be seen as having an impact on the international debate related to reform.



Tuijnman and Postlethwaite (1994), in their *Monitoring the Standards of Education*, note that while the history of large-scale assessment dates back to the early 1960s, a significant development toward a more systematic focus on national monitoring began with the release of reports like *A Nation at Risk: The Imperative for Educational Reform*, the release of the results of IEA's Second International Science Study [ditto], and later, again in the United States, the report from the Conference of the Governors of the 50 states in Charlottesville, Virginia, which sought to frame national goals for education with a strong emphasis on quality. This period of reform was, in part, attributable to the almost universal that the performance of a country's educational system is a key element in establishing its competitive advantage in an increasingly global economy. Education was conceived of as being implicated in a country's economic, social, and personal development and was/is considered one of the main means whereby inequities—social and economic—can be reduced. Perhaps the most dramatic expression of this is contained in the report from the United States, *A Nation at Risk*, in which the authors point to the threat of economic decline as supplanting the past threat of aggressor nations (United States National Commission on Excellence in Education, 1983. Education or, more specifically, the decline in educational standards was cited as the cause of economic decline in the face of intensified global competition. The Commission wrote: "If an unfriendly foreign power had attempted to impose on America the mediocre educational performance that exists today, we might well have viewed it as an act of war. As it stands, we have allowed this to happen ourselves ... we have, in effect, been committing an act of unthinking, unilateral educational disarmament" (p. 5).

Although a model ascribing economic decline to a decline in educational standards is simplistic and likely to be of limited value in addressing or understanding educational and economic policy concerns, the debate engendered by it draws attention to real concerns about educational performance, not only in the United States but also in many OECD countries. It was the concern for excellence— together

with concerns related to equity and efficiency—that gave rise to a greater intensity of focus on education and evidence-based educational policy development.

As education received increased priority in the public policy arena in many countries, it also faced the reality that, as with many other areas of public spending, of real limits to the funding available for educational development. The funding that was available was accompanied by increasing demands for accountability and a better understanding of the relationship between educational expenditure and educational outcomes.

The fullest and perhaps most extreme expression of these concerns is evident in publications like *Reinventing Government* (Osborne & Gaebler, 1993), in which the authors argue for an educational marketplace shaped by the twin imperatives of efficiency and effectiveness. Implicit in their argument is the notion that increased provision and improved instructional quality produce greater numbers of better-prepared students, an outcome that, in turn, results in a more internationally competitive and better-prepared workforce. The role that comparative studies like TIMSS and PIRLS, in particular, can play in such an argument is to direct the focus more narrowly on to assessment of quality in mathematics, science and reading and presumably, therefore, on the production of more productive and high-quality scientists, mathematicians, and engineers in particular and workforce in general.

What is reflected in these kinds of concerns is a shift in focus from managing issues related to the expansion of educational systems in terms of student numbers, to one of managing issues of quality and excellence. In the case of those countries in what might be described as a less advanced stage of educational development, this has meant not surrendering to the imperatives of educational expansion at the expense of considerations of quality. The shift in emphasis from issues of quantity to a greater concern for quality is reflected in the work of IEA.<sup>1</sup>

As we look across the four decades that span the work of IEA, we can observe a progression in terms of complexity. Studies in the 1950s designed to describe one cohort and to compare countries with respect to that cohort gave way to studies designed to address issues of greater complexity. By the early 1980s, changes in design to include multiple cohorts (as with SIMS, for example) allowed for the description of differences among cohorts and, more importantly perhaps, exploration of apparent gains in achievement within and among countries (SIMS with its longitudinal component again provides the pertinent example).

The demand in the early 1990s for regularity of assessments as a means of contributing to knowledge of the quality of outcomes saw the development of TIMSS and the subsequent decision by IEA to make this, together with PIRLS, the core of its work in student assessment. The advent of the TIMSS and later PIRLS assessments has allowed researchers and policymakers to study trends within countries (through study of repeated cross-sections) and, of course, to examine trends among countries.

Table 1 illustrates the way in which IEA's projects address changing demands from policymakers and the information interests of researchers.

What is apparent as we look back over the last 45 years is that as interest in global competitiveness and local accountability has increased so, too, has interest in international comparisons of educational performance. What then has been the significance of these new imperatives on the activities of IEA and what has been their impact?

While judging the impact of any piece of research is not always a simple task—and often results in a rather pessimistic conclusion as to its

---

1. Although policy concerns are significant in shaping the nature of international studies, it should be noted that the studies themselves, because of their organizational structure around expert groups, are informed by the contributions of many influential researchers and educators from around the world. These inputs are also critical in shaping the nature of the international research program.

**Table 1.** Types of Analyses possible with different Study Designs

Description of One Country	Comparisons of Countries
Describe one cohort	Compare countries with respect to a single cohort (pilot study)
Describe differences between or among cohorts (is this difference a gain) (CIVICS)	Compare gains (SIMS, TIMSS)
Study a trend is repeated – cross-sections (TIMSS)	Compare trends (TIMSS, PIRLS, SITES?)

efficacy (Burkhardt & Shoenfeld, 2003)—there is considerable evidence that much of the work of IEA has made, and is continuing to make, a significant impact in the key areas of policy, research, teaching, and curriculum. However, a note of caution: looking for simple relationships between research findings and particular policy outcomes can lead to misunderstandings about the nature of the research-policy linkage. Major policy initiatives or reforms are more likely to be the outcome of a wide variety of inputs and influences, while research is more likely to provide a heuristic for policy intervention or development. For this reason, we have to cast the net wide when looking for evidence of impact.

### Public Visibility

The work of IEA, perhaps more than any other organization, has brought into sharp relief educational outcomes. The extent to which information resulting from IEA studies has entered the public discourse and is visible in the public domain may be one way of assessing impact. This public visibility is arguably a key factor in shaping the public policy debate. While IEA has, in recent years, taken considerable pains to ensure that the results of the research it has conducted are widely disseminated, we might not have anticipated the extent to which this has been achieved.

Over the last 10 years, coordinated media releases have ensured that data release has been maximized. The outcomes of the major IEA studies, including TIMSS, PIRLS, Civic Education, and SITES, have all captured media attention not only in the international media such as Associated Press, CNN, and USA Today, but also in many—if not most—of the major newspapers of participating countries. Internet, too, has provided for a significant increase in the visibility of and accessibility to information resulting from IEA projects and, for that matter, study data. For example, a search on TIMSS reveals some 113,000 citations and no less than 30 educational systems with web sites reporting on their participation in the project. Other web sites reporting on TIMSS include NCTM, the Max Planck Institute, the Mathematics Project Journal, the Washington Post, PBS, and the National Science Foundation.

For each of the major projects other than TIMSS that IEA has sponsored in recent years, PIPLS (with 7,990 citations), Civic Education (with some 6,000 citations), and SITES (more than 50,000 citations), data are reported on country web sites or reported in other public media web sites, including school boards associations or teacher professional organizations. This level of visibility has ensured wide use of the work of IEA.

It is also important to note that dissemination of findings is not restricted to the countries of the OECD. Elley (2002), in reporting on the extent to which results for the 1995 TIMSS study were publicized in low- and middle-income countries, notes that although public release of information was politically difficult in some of these countries, most, if not all, produced national reports that were released and, in some cases, were the focus of intense public scrutiny, for example, through repeated television interviews. Further evidence of the extent of the visibility of the work of IEA is apparent in that, in the United States, the Chairman of the United States Federal Reserve, Alan Greenspan (2004), in his testimony before the Committee on Education and the Workforce, of the United States House of Representatives, appealed to data from TIMSS to express his concern about educationally low per-

forming sectors of the United States population and assumed consequent detrimental impact on the United States economy.

### **Impact on Policy**

Evidence that the work of IEA has entered public discourse and the public policy debate is but a first step in arguing the case for a significant impact. It is also possible to appeal to other sources that argue more strongly for evidence of the impact of IEA's work on educational policy. Earlier I mentioned that the release of data from SISS and subsequently SIMS were influential in shaping the policy debate at least in some of the countries that participated in the studies. It is also evident from the essays provided for *The Impact of TIMSS on the Teaching and Learning of Mathematics and Science by David Robitaille and his colleagues* (2000) that the impact of TIMSS was far reaching following the release of the 1995 TIMSS data. In countries as different as Iceland, Kuwait, New Zealand, Norway, Romania, and South Africa, TIMSS served as a catalyst for curricular review and change. In Iceland, the information collected during the TIMSS study resulted in a recommendation for increased teaching hours for mathematics and science instruction at the elementary level. In New Zealand, the TIMSS results precipitated the establishment of a taskforce with a focus on science and mathematics education. The taskforce was charged with addressing several issues, including the low expectations of success for New Zealand students held by many teachers and parents, underachievement among Maori and Pacific Island students, and teachers' lack of confidence in their ability to teach some aspects of the mathematics and science curricula. In Israel, the headline 'Down in Rankings, Israel Seeks Changes in Education' as reported in *Education Week*, and an associated report of the reforms being implemented in that country largely as a result of what it learned from TIMSS further attests to IEA's impact on educational policy. Further analysis of TIMSS data collected in a manner that would allow identification of

common misconceptions about mathematics and science is also likely to lead to instructional and curricular refinements in many countries.

A publication like *The Impact of TIMSS* that is dedicated to impact provides fairly clear evidence of impact. One could argue, however, that other IEA studies that do not have the benefit of dedicated publications have also contributed in a significant way to influencing educational policy debate.

Findings from Module One of SITES, for example, drew the attention of policymakers to at least three major issues. The first was that in those countries that participated the earlier challenge of providing schools with sufficient numbers of computers had generally been met but a critical gap still existed—that of teachers familiar with and able to use technology for instructional purposes. The second issue was that, with few exceptions, the promise or expectation that computers would transform the curriculum and pedagogy had not been realized. The third issue related centered on the difficulties schools were beginning to have in managing access to the Internet and protecting children from inappropriate materials. The availability of information about the changes that had taken place in terms of the penetration of computers into classrooms also helped change the discourse from questions about, for example, whether or not to invest in computers, to questions about how to facilitate their integration into curriculum and instruction. This study was, I believe, important in showing that while technology has the capacity to shape the teaching learning environment, it exposed the turmoil as schools and teachers sought to integrate technology into the classroom.

Civic Education has also played and continues to play an important role in the educational policy debate.<sup>2</sup> Conducted at a time of significant change in Central and Eastern Europe, this project addressed concerns related to the processes of civic education. For example, how do coun-

---

2. The Civic Education Study continues to stimulate discussion at the OECD about the inclusion of a study in this area as part of INES activities.

tries manage the process of preparing students for citizenship and membership of participatory democracies? This study revealed important differences across and within countries in terms of knowledge about and understanding what is meant by democracy but also of student attitudes towards nation, government, immigrants, and women's political rights. As countries, particularly emergent democracies, struggle to manage the process of developing a sense of nationhood while at the same time allowing for the expression of cultural and ethnic identity, particularly among minority groups, the information from this project will surely continue to contribute to the policy formation process.

### Impact on Research

One of the more obvious ways in which the impact of IEA studies on research can be judged is to examine research publications related to or using the data that IEA has gathered. Of the primary publications of the organization, approximately 11 of study findings were produced up to 1980 but since that time 94 key publications have been produced. The relatively modest output for the period 1956–1980 roughly corresponds to the number of studies undertaken during this time.

A focus only on those publications arising directly from studies in the form of project reports or study summaries seriously underestimates, however, impact along this dimension. Degenhart's *Annotated Bibliography of IEA Publications* (1990), for example, lists approximately 800 publications based on the work of IEA. Unfortunately, no such summary exists for the period 1990 to the present, although some project-specific publications do provide further insights into the impact on research.

For example, Robitaille et al. (2000) in their *Impact of TIMSS on the Teaching and Learning of Mathematics and Science* identify research, particularly research related to educational policy and mathematics and science learning, as an area in which TIMSS has had a particular impact. For some of the participating countries—Macedonia, for



example—TIMSS was the first assessment of student achievement undertaken at a national level. Furthermore, all countries were seen to benefit in some way from the training in large-scale student assessment that was associated with participation in the study. Some direct outcomes of the TIMSS training activities included capacity building within many countries, some of which established dedicated research institutes that used the skills and knowledge developed through the TIMSS experience to enhance their national capacity for assessment and policy development. Also, as evidenced by Robitaille, Beaton and Plomp (2000), TIMSS data have spawned a large number of investigations, both national and international, directed at key policy issues as well as at developing understanding of mathematics and science learning in a more fundamental way.

Keeves (1995) in an earlier review of key research findings from IEA studies identified 11 major such findings arising from the work of IEA. While some of these had been previously identified in the research literature, one in particular (opportunity to learn) was deemed to be one of the early contributions of IEA studies, and it continues to be so. The development of the conceptual model Opportunity to Learn, which explores links between policy, practice, and outcome, has helped shape the design of most, if not all, IEA studies. The evolution and development of this model, which was first proposed in the First Mathematics Study (FIMS) has provided powerful insights into the relationship between the 'intended curriculum' (as prescribed by policy), the 'implemented curriculum' (as reflected in teacher and school practice), and the 'achieved curriculum' (as reflected in student outcomes). It has also provided a framework for increasingly complex study designs and provided researchers with the associated ability of exploring more complex issues related to teaching and learning.

An examination of the transformation that has occurred in the design of IEA studies allows us to observe the transition from the early studies where the focus of the research was primarily descriptive, to those investigations where more sophisticated attempts are made at approx-

imating causal explanations. As a further example of this development, the data from TIMSS 2003 to be released later this year represent the first large-scale international study of student achievement offering data from three cycles of assessment. This achievement represents a significant advancement in study design, scaling, data analysis, and, for that matter, reporting.

Another way in which the contribution to research can be assessed is to consider the extent to which the activities of IEA have been implicated in capacity building. For some countries, participation in studies and training opportunities has significantly enhanced their ability to carry out research and assessment activities. Elley (2002) reports that the National Research Coordinators in the low- to middle-income countries learn much from the training experiences and across a range of activities. He also notes the positive impact that participation has on developing a national assessment capacity.

Even in the more economically advanced countries, centers of excellence have developed in those institutions responsible for the execution of IEA projects. While perhaps already in possession of a group of skilled researchers, these institutions generally have been exposed to new techniques and technologies with the evolving sophistication of IEA studies.

IEA as an organization has also benefited. The ongoing cycle of studies has undoubtedly contributed to the establishment of its centers of excellence, such as the study center at Boston College and the Data Processing Center in Hamburg, amongst others. The papers that are being presented at this conference also illustrate how far research activities related to IEA investigations have moved from simple description and analysis.

### **Impact on Teaching and Curriculum**

As suggested earlier, and as Robitaille et al. (2000) note, TIMSS was a

story “with legs”. While the study’s immediate impact on teaching was less strongly articulated in some areas than others, several countries (notably Australia, Canada, Spain, Japan, and, in particular, the Philippines) used the TIMSS results to bring the teaching of mathematics and science into sharp focus. In the Philippines, for example, government resolve was reflected in a program to train 100,000 mathematics and science teachers over a period of five years.

The impact of TIMSS on the curriculum and teaching is also manifest in other ways. The following examples illustrate how widespread and influential this project has been. In Canada (Ontario), instructional materials were developed as an outcome of TIMSS. In England, the results of TIMSS focused attention on lower achieving students and were also the catalyst for a guidance document that stressed the importance of mental calculations and drew attention to the frequency with which calculators were being used in England. In Iran, the TIMSS results led to major changes to the science curriculum and textbooks. In New Zealand, information from TIMSS contributed to the development of resource materials and professional development programs (designed to address the perceived areas of relative weakness) for mathematics and science teachers.

The impact of TIMSS, however, has not been confined to the outcomes of the assessment data. In the United States (and other countries) the TIMSS Video Study revealed significant differences in teaching practice and in the experience of being a teacher of mathematics. In particular, the United States teachers experienced a much greater degree than teachers in other countries what was described as cultural isolation and had fewer opportunities to work closely with other teachers to improve teaching. These findings informed professional development programs and tools. Like other aspects of TIMSS, the curriculum analysis component of the First TIMSS study has been particularly widespread. The structure and content of the United States approach to curriculum, for example, came under intense scrutiny and debate (see, for example, Schmidt et al., 1999). This scrutiny continues today.

The Civic Education Study has also had an impact on curriculum and teaching. Later at this conference a colleague from Australia will argue that this study has been instructive and influential in facilitating and shaping significant developments in civics and Citizenship Education (CCE) in Australia.

SITES, too, has provided insights into the role that technology might play in teaching and curriculum. The first module of the study identified the gap between the provision of technology and teacher readiness in the use of that technology for instructional purposes. In Norway—a country that has invested significantly in the provision of computer technology—this gap has become a particular concern. Although countries are still absorbing the recently released information from Module Two of SITES, that information is also beginning to have an impact. The case study descriptions from SITES of the wide variety of innovative ways that technology is being used around the world to enhance curriculum and instruction are the subject of ongoing discussions in several countries.

Finally, PIRLS, despite the data release being comparatively recent, has also forced an examination of reading instruction in several countries. The data from the parent questionnaire has highlighted the importance of parental involvement in the reading acquisition process and has raised the question as to how parents might better be involved in their children's learning to read.

### **New Countries, New Responsibilities**

The materials reviewed so far have, in my view, provided ample evidence of the impact of the work on IEA on educational policy, curriculum and teaching, and research. However, most of my comments to this point have focused on the past. It is also important to reflect on the future if IEA, as an organization keen to continue as a leader in international assessment and policy-related comparative education.

Since its inception, IEA has seen considerable growth, not only in its membership but the number of non-member countries that participate in its studies. In the period, 1956 to 1980, three studies were completed including the 12 Country Pilot Survey, The First Mathematics Study, and the Six Subject Survey. Since 1980, IEA has completed 13 projects and currently has four project in various stages of completion. Participation in IEA studies has also changed. As suggested by its title, 12 countries participated in the first IEA project, and, depending on the subject matter being studied, participation in the Six Subject Survey ranged from eight countries for French as a Foreign Language to 11 countries for Civic Education. Later, the Second Mathematics and Second Science studies attracted 23 and 19 countries respectively.

What is notable is that as the number of countries participating in new studies so too did their diversity in terms of educational, social, and economic development. The 12 Country Pilot Study included Poland and Yugoslavia but was primarily a study of educational systems characteristic of high-income countries. With a few exceptions, this pattern remained in effect until the 1980s. The IEA Reading Study in 1990 was the largest international comparative study of education ever undertaken, with more than 30 participating jurisdictions, and it included countries from South America, the Caribbean, and Africa.

The decision to formally establish a regular cycle of studies in mathematics, science, and, later, reading has meant that participation in these two subjects regularly exceeds 40 jurisdictions. The expansion of participating countries has, of course, resulted in the inclusion of many low- and middleincome countries whose social, political, and economic circumstances distinguish them markedly from their OECD counterparts. Language, culture, and orthographic differences have all had an impact on the way in which IEA has come to work. For IEA, the inclusion of the broader range of countries with distinctive local circumstances has meant the development of new ways of working to ensure that all countries can participate and that studies continue to

achieve the highest technical standards.

Diversity also brings, I believe, new responsibilities. While continuing to execute studies that meet the highest technical standard must remain the central focus of the organization, we must also guard against the type of complacency wherein assessment would be the only goal. Our success should not be judged solely on the basis of project completion. Rather, as we assume responsibility for greater numbers of countries with greater cultural and economic diversity, our success should be judged on our capacity and ability to contribute to educational policy reform. For an organization like IEA, the goals must include ensuring that the questions that are addressed in our research and the data that are collected meet the needs of our membership and the larger group of countries that choose to participate in our projects. The pay-off for this expansion surely is large. Those who began the work of IEA spoke of a world educational laboratory, but it is only now that the increased participation of countries from around the world is allowing full realization of this vision. The potential for understanding and exploring the impact of the social context on educational outcomes is far greater today than it was in the early 1960s.

To that end, if we are to increase the impact on research and policy and to raise awareness of the work that we do, we need to carefully examine the various ways in which we can work with participating countries. Training is clearly one area where there is great potential to increase IEA's relevance. To improve the training that is offered, we need to ensure that future training sessions should (and will) focus not only on database use as a technical matter but also on assisting researchers to address more fundamental questions, such as what can (and should be) be answered by data that are collected from IEA studies and how that data can be integrated with other policy relevant information. Future training must also more fully recognize the range of experience and needs represented by the countries participating in our studies.

While the decision to establish our core business around studies of mathematics, science, and reading has ensured continuity and stability for IEA, we need to continue to work in those other areas of activity where we have an established expertise and also to consider new areas of investigation where IEA can make a contribution. The new project in teacher preparation is one example of where IEA's collective expertise and infrastructure can be used effectively. Undoubtedly, IEA will consider other areas in the future.

In summary, I believe there is considerable evidence that IEA has made a significant contribution to key areas of policy, curriculum, teaching, and research. In addition to the evidence I have cited, we can all report other instances of impact in these areas. I am sure that Dick Wolf in his address will provide an additional perspective on these issues.

I can announce that as a means of encouraging young researchers to work with IEA data, the IEA Standing Committee has elevated the status of the Bruce Chopin Award by offering a sum of 500 euros to go along with the certificate of recognition. It is also hoped that the revised review structure will ensure a broader based competition than has been the case recently and that this hopefully will help secure a base of skilled researchers who can contribute to the future of IEA.

In addition, to facilitate further opportunities for secondary analysis of IEA data, the Standing Committee has agreed that the organization should continue to support research conferences of this type, and plans are underway to secure a venue for the next conference in two years time. We look forward to seeing you at our next conference in 2006 and perhaps revisiting some of these issues then.

## References

- Burkhardt, H., & Shoenfeld, A. H. (2003). Improving educational research: Toward a more useful, more influential, and better-funded enterprise. *Educational Researcher*, 32(9), 3-14.

- Degenhart R. E. (Ed.). (1990). *Thirty years of international research: An annotated bibliography of IEA publications (1960–1990)*. The Hague: International Association for the Evaluation of Educational Achievement.
- Down in rankings, Israel seeks changes in education. (2004, 3 March). *Education Week*. Retrieved from <http://www.edweek.org>
- Elley, W. B. (2002). *Evaluating the impact of TIMSS- R (1999) in low and middle income countries*. Report for the World Bank, Christchurch, New Zealand.
- Foshay, A. W., Thorndike, R. L., Hoytat, F., Pidgeon, D. A., & Walker D. A. (1962). *Educational achievement of thirteen-year-olds*. Hamburg: UNESCO Institute for Education.
- Greenspan, A. (2004). *Address to the Committee on Education and the Workforce, United States House of Representatives March 11, 2004*.
- Keeves, J. P. (1995). *The world of school learning: Selected key findings from 35 years of IEA Research*. The Hague: International Association for the Evaluation of Educational Achievement.
- Osborne, D., & Gaebler, T. (1993). *Reinventing government: How the entrepreneurial spirit is transforming the public sector*. New York: Plume.
- Pelgrum, W. J., & Anderson, R. E. (1999). *ICT and the emerging paradigm for life long learning*. Amsterdam: International Association for the Evaluation of Educational Achievement.
- Postlethwaite, T. N. (1974). Introduction. *Comparative Education Review*, 18 (2), 157–163.
- Robitaille, D. F., & Beaton A. E. (2002). *Secondary analysis of the TIMSS data*. Dordrecht: Kluwer Academic Publishers.
- Robitaille D. F., Beaton, A.E., & Plomp T. (2000). *The impact of TIMSS on the teaching and learning of mathematics and science*. Vancouver: Pacific Educational Press.
- Schmidt, W.H., McKnight, C.C., Cogan, L., Jakwerth, P., & Houang, R. (1999). *Facing the consequences*. Dordrecht, The Netherlands: Kluwer Academic Publishers.



- The World Bank. (1999). Education sector strategy. Washington D.C.: The International Bank for Reconstruction and Redevelopment World Bank. Report # 79.
- Torney-Purta, J., Lehmann, R., Oswald, H., & Schulz, W. (2001). *Citizenship and education in twenty-eight countries: Civic knowledge and engagement at age fourteen*. Amsterdam: International Association for the Evaluation of Educational Achievement.
- Tuijnman, A. C., & Postlethwaite, N. T (1994). *Monitoring the standards of education*. Oxford: Pergamon.

## CHAPTER 12

# **The Contribution of IEA to Research and Education\***

*Richard M. Wolf†*

I want to thank the organizers of this conference for their kind invitation to me to present what I consider the important contributions that IEA has made to both research and education over the years. I have been a part of IEA and IEA has been a part of my and my family's life for over forty years now. It has been an exciting ride, to say the least. Despite my continuous connection with IEA over this period, trying to identify its contributions is no easy task. The reason for this difficulty stems from the fact that isolating the contribution of a specific activity or organization to some larger social structure such as education is risky. One is not sure how much to ascribe to the specific activity and how much to other factors in a society. What I am offering here are my best guesses as to what IEA has contributed to research and education. It is not inconsiderable, I believe, and I shall try to demonstrate this in this paper.

I would like to begin by referring to the 1950s. There was a great deal of ferment in education at that time. As one example, two books were published in the United States that became best sellers with the general public. The first was published in 1955 by Rudolph Flesch and was titled, "Why Johnny Can't Read" and the second was published in 1958 by Admiral Hyman Rickover and was titled, "Swiss Schools are Better Than Ours". Besides being best sellers, these two books had two other things in common. First, each was highly critical of primary and secondary schooling in the United States. Second, neither book presented any systematically gathered evidence to support the allegations

---

\* Keynote address at the first IEA International Research Conference (IRC-2004).

that were made. All one had to rely on were some scattered anecdotes throughout each tome. (Incidentally, it is not at all clear how a United States Navy admiral could become an expert on the educational system of a totally land-locked country). The second source of ferment in education was the launching of Sputnik I in 1957. It raised a number of questions about the state of American education.

Soon after the publication of these books, the launch of Sputnik I, and other similar books and reports in other countries, a group of educators began to meet in 1955. One product of that early meeting was a book on evaluation by F. Hoytat of Belgium. William Wall who was the Director of the National Foundation for Educational Research in England and Wales and the first chair of IEA (1958-1962) pushed the group to conduct a pilot study to test the feasibility of conducting multi-national studies of educational performance. Performance. The first formal meeting of the group was held in Eltham, England in 1958. Researchers attended it from England, the United States and several other countries. The group began to discuss ways in which research could contribute to the improvement of education not only in their countries but in other countries as well. It was at that meeting that the proposal that an exploratory study be undertaken to determine the feasibility of carrying out an empirical study of student performance in various school subjects was considered and decided.

Several work groups were formed to develop short tests in several subjects that could be administered to students (in their own language) to measure their performance in these subjects. Since there was no real funding for the study, it was decided that a representative from each country would take responsibility for translating the tests into their own native language, identify small judgment samples of students who would take the tests, administer them, score them and bring the results back to the main group. While the researchers were well aware of the limitations of their study, their prime concern was in testing the feasibility of such a study and how well researchers from differing systems of education could work together. It was also hoped

that the researchers could begin to learn something about the educational systems of the other countries.

While the study was actually a feasibility exercise, there were two notions that were of concern to the researchers: (1) what could countries learn about their own educational system from such an exercise, and (2) what are the antecedents and correlates of achievement and how similar were these relationships across countries. A third notion that emerged a bit later was that there were certain variables that could only be tested in a multi-national study, e.g., the influence of age of starting school on subsequent performance.

Some countries began schooling at age 5, others at age 6 and others at age 7. Testing the influence of these different starting ages on later achievement could not be carried out within the borders of a single country. It required a multi-national study. The ultimate goal of the work was to identify ways in which education could be improved throughout the world.

The group called itself IEA beginning in 1960 but was not legally incorporated until 1966. It carried out the study, which had been designed, and instruments developed over a couple of days. As a piece of research, it would never win any awards for scientific rigor, but as a feasibility study, it was successful enough to encourage the researchers to undertake what the English would call "a proper study".

The researchers decided to conduct a study in the area of mathematics, believing that this would be the easiest subject in which to test because of the use of numbers instead of relying heavily on language. While the investigators were probably correct in their choice of mathematics, there were still enormous problems to overcome. A few additional educational systems were invited to participate and the first large-scale study of mathematics achievement was begun. (I use the term "educational systems" instead of countries because from the very outset of IEA some countries had two different systems of education, notably French and Flemish Belgium and England and Scotland in the United Kingdom).

It was decided that, along with the mathematics tests, attitude scales for students and questionnaires for students, teachers and school heads would be developed and administered. The levels to be tested were: (1) 13 year olds, (2) students in the eighth grade of school (largely 13 year olds), and (3) two groups of students in the last year of secondary school, students specializing in mathematics and students not specializing in mathematics. Countries were given the option of testing a group midway between these two grade levels. Furthermore, it was decided that probability sampling would be undertaken to obtain representative samples of each defined population and the samples should be large enough for sufficiently precise populations estimates and for statistical testing.

The researchers were strongly committed to carrying out a high quality research study. However, few, if any, of the researchers realized how large and complex a job they had taken on. This was only slowly discovered over the next three years. The fact that the study was carried out and completed as well as it was is due to the heroic efforts of a number of people. The report of the study received major coverage by the press all over the world. For the very first time there was evidence on which to make judgements and improvements in education rather than relying on mythology, unsupported allegations, and the whims of educational or political officials. This is perhaps the major contribution that IEA has made to research and education. Although there have been many misguided attempt to use evidence to make educational decisions, the notion that evidence should be used in making educational decisions is well established. This, however, is not IEA's only contribution to research and education. Let me cite what I regard as three other major contributions and some minor, but I believe important ones.

The three major contributions are easily remembered since they all begin with the letter "C". They are competence, cooperation, and computers. First, there is competence. While IEA began as an informal association of researchers (IEA was not formally incorporated until

1966). It was able to enlist an impressive roster of specialists to work on IEA studies. The group included, among others, Ralph Tyler, John Tukey, Julian Stanley, James Coleman, John B. Carroll, Sten Henrysson and Herman Wold. As researchers, they had rather similar training and backgrounds that easily overcame differing national views. In addition to the IEA researchers, IEA was able to draw on many of the best minds in the world in the areas of education, psychology, sociology, economics, political science, and statistics as consultants. These specialists were quite happy to work with IEA, usually for no payment. The studies that IEA was planning or carrying out engaged their interest and IEA has benefited greatly from their expertise. This was over and above the expertise that already existed in IEA which was considerable. The initial research workers in IEA included Benjamin Bloom, Robert Thorndike, C. Arnold Anderson, Douglas Pidgeon, Gilbert Peaker, David Walker, Torsten Husen, Sven Hilding, and Neville Postlethwaite among others.

The second major contribution of IEA was the promotion of international cooperation. While nations had differences, this did not spill over into IEA work. I believe that the reason for this is that the IEA members in the 1960s and 1970s were, for the most part, researchers and were working towards a common set of purposes. They had similar training and spoke a common language, the language of science and research. This is not to say that there weren't disagreements. There were, in fact, many of them from the very outset.

Two examples that spring to mind are: (1) the discussion over the use of multiple-choice vs. open-ended questions in the mathematics tests for the first major IEA study, and the use of: (1) analysis of variance or (2) multiple regression analysis for data analysis. It wasn't until later that we found out that the two are mathematically identical. However, both were actually used in the data analysis of the first mathematics study. Another issue that continued to vex the IEA researchers over a substantial period of time was whether to define populations of students in terms of age or grade in school. There are good reasons to sup-

port either approach but IEA has found a way to deal with this by testing samples in adjacent grades that contain the great majority of an age group. Data can then be analyzed either by age or grade.

The third major contribution of IEA has been in the area of what can loosely be described as computers but includes considerably more than hardware. The first IEA mathematics study tested over 250,000 students in over 10,000 schools in twelve different educational systems. Analyzing that amount of data would be impossible without high speed, high capacity computers. It was a stroke of good fortune that such computers became generally available just as IEA was launching its first mathematics study. In addition, optical mark readers also became commercially available at the same time. It was thus possible to design a single answer sheet for each tested population and that could be used in all twelve participating systems of education. The answer sheet not only contained response areas for multiple choice mathematics items, but room for open-ended responses and responses to sixty-five attitudinal items. Questionnaires for students, teachers, and school principals were asked in paper and pencil form and had to be coded at each national center, entered onto punch cards and transmitted to the international data center for entry into computer.

IEA's use of the best computer hardware available should not mask the fact that the software developed for IEA studies is clearly at the forefront of research. The Data Entry Manager, developed for the IEA Reading Literacy Study of the early 1990s, is one of the keys to the success of all IEA studies since that time. As a testament to the success of IEA developed software, OECD's WEI and PISA projects make use of some of this software along with UNESCO's SACMEQ project.

IEA today is a far different organization than it was when its first began. It is an organization that is incorporated under Belgian law, has over 70 member systems of education, has a headquarters in Amsterdam with a permanent staff, produces newsletters and bulletins and has a process for pre-publication review of study reports.

This last item arose from the fact that early IEA study reports were strongly criticized for being written in “researchese” and almost unreadable to all but a few researchers. IEA has changed considerably from its earliest days in terms of the composition of its General Assembly, the main governing body of IEA. As previously noted, IEA originally consisted of a group of researchers. Currently, the IEA General Assembly consists largely of officials of ministries of education along with some researchers. This is good in the sense that it improves the chances for funding IEA studies and the organization as a whole through the payment of dues. Having officials from ministries of education on the General Assembly also provides a way for the various systems of education to decide what studies would be most useful to their work and what questions they want answered. The drawback to this is that some of the ministry officials sitting as IEA General Assembly members are not in a good position to judge research proposals and have to rely on the expertise of IEA staff and other consultants to ensure good research. So far, this has worked rather well.

IEA has conducted a substantial number of studies over the past forty years. Some have had considerable impact internationally while others have had great impact in some countries but not in others. In viewing all IEA studies, it seems that one of the major outcomes of IEA research work is their contribution to what Torsten Husen, the second Chair of IEA, refers to as the “massification” of education. Prior to the IEA studies, a widely held view in many countries was that the pool of talent was limited and a selection system was needed to identify those students who were considered capable of completing a secondary education. There was a deep-rooted view in a number of countries where an economics of scarcity gave rise to a psychology of scarcity. According to this view, the pool of talent was regarded as being limited (quite limited in the view of some countries). The IEA studies of the 1960s through the 1990s helped destroy this myth. Those studies showed that many countries had underestimated their talent pool. IEA studies showed that many more students could be educated to the



completion of secondary school without causing any harm to the performance of the highest ability. In other words, the yield of an educational system could be increased markedly by retaining a much greater percentage of an age group in school. To see evidence of this "massification" of education, one only has to turn to enrolment figures in a number of countries. The percentage of an age group completing a secondary education has increased markedly over the past 30 years, as shown in the IEA studies, with no loss in the achievement levels of the highest level of students.

Another major contribution of IEA studies was the development and use of measures of student achievement as outcomes of an educational system instead of previous work that usually relied on just enrollment figures. Furthermore, IEA introduced the notion of measuring achievement at different points in an educational system instead of just at a single point, usually the end of secondary education. This was a distinctly novel contribution since it allowed a country to track achievement throughout the levels of an educational system.

On the other hand, despite forty years of effort, IEA has not succeeded in developing suitable ways of measuring student achievement at the end of secondary school for comparative purposes. IEA has finally learned that there are so many differences in the various systems of education that any attempt at comparison is doomed to fail. The major differences between systems are: (1) the age of students in the last year of secondary education (the range is from age 16 to age 20), (2) the percentage of students in an age group still enrolled in school (over the years this has shown a range from 9% to over 80%), and (3) the amount of specialization that occurs in a number of systems (in science, for example, some students at the end of secondary school may have taken only one course in physics while in other educational systems student may have taken four or five courses in physics). Reluctantly, IEA has had to confine its studies to measuring student achievement during the years of compulsory education. This is still a notable accomplishment.

Sometimes overlooked is IEA's development of tests that could be used fairly in a large number of educational systems. This involved the development of a newer technology of testing, one with an international dimension. The use of new technologies in testing such as item response methods, notably Rasch scaling, and imputation procedures that allow one to estimate a student's performance, despite the fact that a student takes only a limited number of items, have worked well. These procedures have allowed for a common test to be used in large scale surveys, the use of a much larger pool of test questions to be used to cover a domain, and ways of linking tests at different levels within a study and with tests used in previous IEA studies. This allows for achievement comparisons over time. The recent repeat of the IEA Reading Literacy Study of the early 1990s is a case in point as well as the TIMSS Repeat study.

IEA has also done pioneering work in the use of complex sampling procedures in the conduct of its studies. These procedures have allowed IEA researchers to obtain unbiased estimates of student achievement and other variables with a high level of precision at a reasonable cost. Subsequent research projects such as PISA and SACMEQ have benefited from IEA work in this area. A notable contributor to this work was the late Gilbert Peaker.

The idea of using research results to make recommendations about systems of education is hardly new. However, the IEA studies did provide an impetus to using such results since they contained pertinent information about other systems of education that could be used for comparison purposes.

Going further, some educational systems have used item analysis results from IEA tests to provide information to curriculum centers so that modifications could be made to curricula. Hungary has been a leader in carrying out this type of work.

IEA provided the first empirical evidence in support of C. P. Snow's construct of a two-culture theory. In its six-subject survey of the 1970s,

IEA tested the same students in both science and literature at the 14 year old level and at the end of secondary school. This provided a way of testing Snow's theory. The evidence clearly supported Snow's theory in the countries where such a phenomenon was hypothesized to exist and showed that it was not a universal phenomenon since several countries showed no virtually no evidence of the existence of two cultures, one scientific and one literary. However, in education, tradition frequently takes precedence over evidence. Countries that have a highly developed two cultures system continue to maintain it, despite Snow's warnings that it can hurt a society in the long run.

A contribution of IEA to educational research that has often been overlooked is its training function. The seminars held in Granna, Sweden during the 1970s has been the most visible aspect of IEA's training work. Probably more important, however, has been the hands-on training provided to national research coordinators as part of IEA's studies. The research coordinators who are charged with carrying out the work of an IEA study in their country often come to the task with limited training and experience. This was recognized early on and training sessions have been provided at each research coordinator's meeting so that they would be able to carry out the work of the study when they returned home. This training has increased the research capacity in many countries, often under the leadership of these research coordinators. Some have risen to prominent positions in research in their countries. An argument could be made that this may be IEA greatest contribution to research in education.

Identifying IEA's substantive contributions to education is not an easy task for two reasons. First, there have been so many detailed results from IEA studies that it is impossible to catalog them all. Second, isolating the contribution of results from IEA studies from other research studies is simply impossible. The best thing that can be said with regard to this is that when the results of IEA studies are in agreement with the results of other studies, they add to the strength of the results and their likely acceptance by the education community. In some

cases, however, IEA results have had a considerable influence on educational policy and practice.

While the identification of specific IEA results to education is extremely difficult, if not impossible, some broad contributions can be identified. From the very first study in mathematics, IEA researchers have sought to determine the relationships between out of school variables such as home and community background with achievement. These relationships have been found to be rather substantial. Subsequent studies have confirmed this finding although the level of relationship varies across school subjects, ages and systems of education. The relationships are generally the highest in the area of reading and somewhat lower in mathematics. They are still significant and meaningful, however.

The importance of out-of-school factors is still being studied. In the TIMSS study, students were asked to report on whether they received extra instruction in mathematics and science (outside of the regular school program). This phenomenon has received different names from different investigators. These include extra tuition, extra-class instruction, and shadow education. It is only now starting to be studied in a systematic and detailed way although it has been going on for years in various places. In Japan, it is called the Juku but has different names in different countries. For example, Japan has carried out several impressive studies in this area. We are only beginning to understand this phenomenon and its possible effect on achievement. We do know that from about 20% of students to a very high percentage of students receive extra school instruction, often at considerable cost to their families. In general, lower performing students report a higher level of such extra-school instruction than higher performing students. This has only been studied in mathematics and science so far, but the study of extra-school instruction can be expanded considerably to increase the number of subjects and frequency of such instruction. As noted previously, Japan has already conducted some interesting studies in this area but not as a result of IEA research.

IEA has added considerably to our knowledge about gender differences in achievement. Unfortunately, the picture that emerges is not that all clear cut. In some subjects at some age or grade levels, there are considerable differences in the achievement of boys and girls while in other subjects there are virtually no differences. Furthermore the differences are often inconsistent. Also, in some educational systems, there are boy/girl differences in achievement but not in all countries. It is a somewhat confusing picture. However, it does demonstrate that there are no inherent differences between boys and girls achievements in different school at different age or grade levels.

One possible explanation for these mixed results that has been put forward for gender differences in some subjects at some age or grade levels but not in others is that where a school subject is learned individually, e.g., at one's desk, especially at younger levels, gender differences are small or non-existent. This appears to hold for both reading and mathematics. However, when students at these lower levels work in groups, boys tend to outperform girls. In mixed groups, boys tend to take over the work from the girls, often pushing them aside or giving them menial tasks to perform. This was quite noticeable in the first computers in education stud where students worked in groups around a single computer. It is also noticeable, to some extent in science studies that involve laboratory work. To be sure, such behavior can stem from defined gender roles in different countries. Despite analysis of gender differences in virtually every study IEA has conducted over the past forty years, the issue of gender differences (or the lack of them in a number of instances) has not been adequately explained. We do know, from all the conflicting evidence, that gender differences in achievement appear to be more a cultural phenomena than a biological one.

While IEA was never regarded as an educational Olympics, the league tables of achievement that have resulted from IEA studies have received widespread attention; perhaps more attention than anything else in the IEA studies. These results have been interpreted (and mis-

interpreted) in many places. Poor results have often been a spur to reform an entire educational system. In the United States, for example, the results of the IEA six subject survey served as a basis for a call to reform U.S. education in the 1983 government publication, "A Nation at Risk".

Finally, consider for a moment the difference between the 1950s when all sorts of pronouncements and allegations were made about the state of education in many countries without any supporting evidence and the 1980s onwards where people have come to expect evidence to support pronouncements about the state of an educational system. I believe that IEA has contributed substantially to this change. In fact, it may be IEA's great contribution to education. One hopes that the educative function of IEA will continue to guide the process of basing educational policies on a foundation of evidence and not on whims or simple beliefs that have no evidential base.





PART C

**REFLECTIONS ON CONTRIBUTIONS OF  
IEA RESEARCH FROM THE PERSPECTIVE  
OF EDUCATIONAL SYSTEMS**





## CHAPTER 13

# The Involvement of African Countries in the IEA Studies over 50 Years

*Sarah Howie*

*In this chapter, I reflect on the IEA's presence in and impact on a continent and a country. First I will present an overview of Africa's participation in IEA studies, and thereafter I will discuss some of the footprints that the IEA has left on the African continent with a focus on South Africa. Finally, some conclusions and reflections are presented. But I will start off by reflecting on how I got to know the IEA and how I subsequently became involved in a number of IEA studies beginning in 1995.*

### Background

In May 1995, I joined the South African Human Sciences Research Council (HSRC) and by July I was on my way as a young twenty-something-year-old to my first National Research Coordinators' (NRC) meeting for TIMSS (1995) in Vancouver, Canada, on behalf of the South African NRC Derek Gray and the HSRC. I was still studying for my Masters' degree and had not yet absorbed all the salient points of the project and was totally overawed by the prospect of representing South Africa's participation in the study. I arrived in a grey, overcast city with no luggage, and until 4:45pm on the Sunday before the meeting started, I faced the prospect of going to the cocktail reception (and the meeting) in my travel-stained clothes – and then my luggage arrived.

My first impressions of that meeting were of a large group of highly educated, very smart, well-organized and much older people. Al Beaton was the International Coordinator at the time, Tjeerd Plomp, was the then Chair of the IEA, and David Robitaille, Bill Schmidt and Bob Garden were all in attendance at that meeting. The latter I remember well because of our jovial, but competitive conversations about the

rugby teams of our countries, South Africa's Springboks and New Zealand's All Blacks, especially since South Africa would go on to win the World Cup that year and Nelson Mandela made an appearance at the World Cup highlighting the theme of reconciliation nationally and internationally.

My first encounters at the meeting revealed the extent of the collegiality and friendship among country representatives and the IEA team, and this remains for me one of the hallmarks of the IEA. The Nordic team (colleagues from Norway and Sweden as well as the Netherlands and Germany) were very much in evidence, and not only were they essential in the intellectual contribution of the 'Viking Rubric' (the partial credit model) scheme adopted by TIMSS, but they were also the life and soul of the socials and parties in those years. In the years since this meeting, I have so often heard others refer to the IEA 'family' and reiterate how much people appreciate the warmth of the individuals in the organization in addition to their professionalism and technical expertise. Furthermore, I experienced firsthand the helpfulness of the IEA partners from Statistics Canada, Boston College and the Data Processing Centre. The success of South Africa's first study (TIMSS 1995) was largely due to the exceptional assistance of Jean Dumais and Pierre Foy of STATCAN, Ina Mullis, Mick Martin, Eugene Gonzalez, and Teresa Smith at Boston College, and Heiko Sibberns from DPC. They (and others from those institutions) went on to assist the South African team with great expertise and patience in subsequent studies.

My overarching impression of the international coordinating group was of a highly professional team; this made me realize that I and the South African team had an awful lot to learn about running the study nationally. This meeting was the start of what was to be an incredibly steep learning curve for me personally, but also for South Africa, in the field of large-scale assessment and international comparative assessments. It was also the start of some long and enduring friendships and little did I know then, also a path to my doctorate-- obtained in 2002 at the University of Twente under the supervision of Prof Tjeerd Plomp.

This was also the year of the birth of the Centre for Evaluation and Assessment at the University of Pretoria. After that first meeting, I went to many subsequent IEA meetings: for TIMSS 1995, TIMSS 1999, SITES M1 and M3, PIRLS 2006, and now PIRLS 2011, mostly in the capacity of NRC. After nearly 17 years working with the IEA, involving six studies (on a country level), visits to many countries and exposure to so many cultures, my appreciation of the organization, its leaders, partners and country representatives has grown. The learning opportunities provided by the meetings and training sessions have been remarkable and I have seen firsthand how these translate into building research and assessment capacity in a developing environment. I have seen the IEA under the past three Chairs, all remarkable men in their own right and all of whom have made a substantive contribution to the organization in various ways. I continue to respect and am still in awe of the amazing organizational ability and expertise of the IEA Secretariat, the International Study Center at Boston College, Statistics Canada, the IEA Data Processing Centre in Hamburg and that of their partners in all the studies I have had the privilege to be involved with.

## Introduction

International studies measuring educational achievement have a number of purposes. Internationally, these studies are of interest to Ministries of Education and international agencies such as the World Bank, UNESCO and OECD. Much of this interest lies in comparing levels of national achievement among countries and examining differences among countries in order to identify major determinants of national achievement with particular interest in malleable factors. The findings of many IEA studies have had an impact on policy and practices across a number of countries.

Aside from the five purposes of international comparative studies (i.e., mirror, monitoring, decision-making, research and enlightenment)

previously reported in Plomp, Howie and McGaw (2003) and Howie and Plomp (2005a), there are some additional benefits of participating in this type of study that pertain more to developing or less developed countries (Plomp, Howie & McGaw, 2003) including those in Africa. These additional benefits may be divided into four areas (Howie, 2000).

First, international studies (e.g., IEA and SACMEQ<sup>1</sup>) have been found to contribute substantially to the *development of research capacity* in developing countries, and this has been my personal experience in the work I have done with the IEA. TIMSS significantly developed the capacity in South Africa (and in other countries) to undertake large-scale surveys involving assessment, which PIRLS 2006 in particular has built on and which provided the vehicle for many Masters and Doctoral research. South African researchers were introduced to the latest research methods and offered substantial training and assistance in their use throughout the research process.

Second, these studies present an opportunity for developing countries to collect *baseline data* in certain subject areas, where previously there was a vacuum. This was evident in South Africa for the TIMSS, SITES and PIRLS studies; in Morocco, Tunisia and Botswana for TIMSS; and for PIRLS also in Botswana.

Third, establishing a national baseline through an international study *heightens awareness of what other countries* around the world are doing, and points to lessons that can be drawn from them. For example, TIMSS was the first international educational research project in which South Africa participated after years of political and academic isolation, providing the first opportunity to review and compare data from

---

1. SACMEQ, or the Southern African Consortium for Monitoring Educational Quality, was initiated in Zimbabwe in the mid-1990s and involved countries in southern and eastern Africa. It was established by the International Institute for Educational Planning of the United Nations Educational Scientific and Cultural Organisation (UNESCO), and to date has conducted three studies in Africa.

South Africa with those of other countries. The disappointing result of this comparison led the then Minister of Education to announce, during a parliamentary debate, that his department would review the data in order to design new curricula to be introduced by 2005. Later in the chapter, I discuss how the impact of PIRLS 2006 revealed the effects of this awareness in South Africa (Howie & Venter, in press). This was also the experience of Botswana (Moahi, 1992).

Finally, education jostles for attention with many other needs in developing countries (e.g., health, poverty, HIV/Aids, rural development). In this context, the fact that the results of international and not merely national studies are available assists researchers, practitioners, and policy makers to *highlight priorities* (Plomp, Howie & McGaw, 2003).

### **Involvement of African Countries in IEA Studies**

In general, there has been a substantial increase in the number of African countries involved in national and international assessments of the educational achievement of their students, especially since the 1990s. In the past decade many developing countries have carried out national assessments and more have participated in international comparative studies, previously confined mainly to industrialised countries.

Twelve African countries have participated in 15 IEA studies to date (see Table 1). A few countries have only participated in one study so far (Algeria, Ivory Coast, Kenya, Swaziland). Algeria is a relative newcomer to the IEA studies, with its first study being TIMSS 2007. Ivory Coast, Kenya and Swaziland only participated once each in the 1970s. Ivory Coast and Swaziland participated in the Second International Mathematics Study, whilst Kenya took part in the Written Composition Study. Other countries such as Morocco, Nigeria and South Africa (Howie, 1997; Howie, 2001; Howie & Hughes, 1998; Howie, 2004; Howie & Pietersen, 2000; Howie, Venter & van Staden, 2008; Howie, Venter & van Staden et al., 2009) have participated in

four or more studies. Nigeria participated in the earliest studies and seems to have withdrawn since 1992. Despite Zimbabwe's intention to participate in later studies, it only participated in the Second International Science Study and the Reading Literacy Study 1991.

The last decade has seen a significant increase in the number of studies; there has also been an increase in the number of African countries participating in international studies. South Africa was the only African country to participate in TIMSS 1995 (Howie, 1997); in 1999, the country was joined by Morocco and Tunisia (Howie, 2001); in 2003, Botswana, Egypt and Ghana joined South Africa, Morocco and Tunisia; and, in 2007, South Africa did not participate but Algeria, Botswana, Egypt, Ghana, and Tunisia did. South Africa participated in the second study of the Progress in International Reading Literacy studies, PIRLS 2006 (Howie et al., 2009). It was one of only two African countries in this study (the other being Morocco, which previously was the only African country participating in the PIRLS 2001). South Africa's participation as the only African country was important as most of the participating countries, besides the USA and Canada, were European, with only two countries in Asia and South America (Howie, Venter & van Staden et al., 2009). A similar situation arose with the SITES (Howie, 2010) as South Africa was the only African country to take part in all three studies.

To date, the most popular study for African countries was the Trends in Mathematics and Science Studies and in 2007 five countries participated (Algeria, Botswana, Egypt, Ghana and Tunisia). Ghana had previously taken part in the Second Science Study in the 1980s and did so again in TIMSS 2003. South Africa has participated in four of the five TIMSS studies (1995, 1999 and 2003, 2011) and two of three Progress in International Reading Literacy Studies (PIRLS, 2006, PIRLS 2011), both of which are international achievement studies. South Africa also participated in all three Second International Technology in Education Studies (SITES M1, M2 and 2006), which were not achievement-based studies.





## IEA Footprints on the African Continent

For the most part, the results of the African countries were rather low compared to the international average of the studies in which they participated. Newspapers often highlighted comparisons with the results of specific countries and in particular other African countries or developing environments. The findings were often surprising in that achievement results were unexpectedly low in mathematics, science and reading, and considered disappointing in the national context where no known benchmarks were in place. There were also other surprising findings, such as the diversity in the number of days of schooling or the large classes in many African countries. For most African countries, national assessments are not yet in place, and prior to the advent of the IEA and, later, the Southern African Consortium for Monitoring Educational Quality (SACMEQ I,II,II), there was few objective data on educational achievement. The disappointing findings served as a wake-up call in many cases and sometimes provided an opportunity for reform. In the TIMSS study of 1995, one NRC referred to the results as the most controversial issue in education in the last two decades (Robitaille, Beaton & Plomp, 2000). In another case, the President of the country called for a rescue package after some disappointing rankings (Elley, 2005, p.4). In Botswana, data that the Education Commission collected using the IEA Six Study Survey in 1976 provided the basis for a government policy on education that expressed a commitment to using educational research and evaluation in seeking solutions to educational problems (Moahi, 1992).

Apart from the surprising and disappointing achievement results, the rich contextual data provided invaluable information for policy makers and researchers about education systems, schools, classrooms and student background. This addressed in a significant way the criticisms of achievement survey results (INEADE, PASC, CONFEMEN; Naumann, 2005) in parts of Africa that failed to take into account student backgrounds and circumstances such as access to home language

nourishment, living conditions and socio-economic status.

Despite the reactions within the African countries and widespread publicity, few researchers published reports on this reaction and/or impact. In a number of African countries, national reports are written and published locally. However, there is very little literature on African country experiences in IEA studies available internationally, and, to my knowledge, there is only one article (Moahi, 1992) outside South Africa that discusses the impact of IEA studies within an African country. Therefore in the next section, the case of South Africa is elaborated.

### **IEA Footprints in South Africa**

South Africa covers an area of 1,220,088 square kilometres and is located at the southern tip of the African continent where it is bordered by Namibia, Botswana, Zimbabwe and Mozambique. Following the first democratic elections in 1994, the country's borders were redefined into nine provinces. The population of South Africa is approximately 48 million and is characterised by considerable disparities in the distribution of wealth, living conditions and opportunities for advancement, and this variance is visible across provinces. Compared to the rest of Africa, South Africa is regarded as a strong economy. It is classified by the World Bank as an upper middle income country.<sup>4</sup>

South Africa is a multicultural and a multilingual society. There are currently 11 official languages designated by the constitution of South Africa. Until 1993, English and Afrikaans were the only two official languages in South Africa. In 1993 the new South African constitution, formed as a result of the dissolution of the Apartheid government,

---

4. South Africa is considerably ahead of most countries in Africa with a GNP per capita of US\$ 3690, yet 34% of South Africans live on less than US\$2 per day (UNESCO , 2007, p. 229) and a large proportion of this group is located in the rural areas.

designated that the nine most prominent African or indigenous languages be incorporated as joint official languages together with English and Afrikaans. The most commonly spoken language is isiZulu (24%) followed by isiXhosa (18%), Afrikaans (13%). The remaining languages are English (spoken by less than 10%), isiNdebele, Sepedi, Sesotho, Setswana, SiSwati, Tshivenda, and Xitsonga (South Africa Info, 2006).

The South African constitution of 1996 makes a commitment to provide for the right of all children to be educated in their own language. Accordingly, the Department of Education's Language-in-Education Policy (DoE, 2007) recommends that the learner's mother tongue be used for teaching and learning wherever possible, especially in the foundation phase, grades R(eception) to 3. However, this recommendation is not uniformly implemented nor is it currently standard practice that every learner is able to be educated in his/her first language. The majority of schools offer schooling in grades 1-3 in one of the African languages (the remaining in English and Afrikaans), but nonetheless many African students do not receive instruction in their first language. This situation is further complicated, as at the end of the foundation phase (grade 3) the current Language-in-Education Policy (DoE, 1997).<sup>5</sup>

Education in South Africa is compulsory and free for grades 1 to 9, and non-compulsory for grades 10 to 12. South African schooling is divided into four phases: the foundation phase (grades 1 to 3), the intermediate phase (grades 4 to 6), the senior phase (grades 7 to 9) and the further education and training phase (grades 10 to 12). Primary school spans grades 1 to 7, whilst grades 8 to 12 constitute the secondary school. There are about 12 million learners enrolled in schools in South Africa. In general, teachers in government schools are faced

---

5. The policy mandates a switch from the use of the mother tongue of learners, as the language medium of educational instruction, to English as the language of teaching and learning.

with large classes; the PIRLS and TIMSS studies revealed the average class size as greater than 45 and sometimes as large as 120 (Howie, 2003a, Howie, 2003b; ; Howie, 2005a; Howie, 2005b; Howie & Plomp, 2002; Howie, Marsh & Hughes, 2000; Howie, Scherman & Venter, 2008).

### *South Africa's Entry into the IEA Studies*

The focus on evaluation, assessment and monitoring is still relatively new in South Africa and has received much more attention in the last decade than previously. The Third International Mathematics and Science Study (TIMSS 1995) was conducted in 1995 in South Africa. South Africa went on to participate in other TIMSS studies, as well as the Second International Technology in Education Study (SITES) and the Progress in International Reading Literacy Study (PIRLS) 2006 (see Table 1). The country also participated in regional studies run by the Southern African Consortium for the Monitoring of Educational Quality (SACMEQ) I and II. In 2000, the National Department of Education introduced South Africa's own national assessments, the Systemic Evaluations at grades 3, 6 and 9 (sample-based) and most recently, in 2011, introduced the Annual National Assessments for grades 1-6 and grade 9 (a census).

Emerging from decades of political and social isolation and a legacy of inequity, in 1994 South Africa had no systemic overview of the schooling system with regard to education quality (Howie, 2002). Any evaluation of education or research was focused within an individual provincial education department. With the dramatic expansion of the education system in 1994 focused on access to education, there was an urgent need to ascertain the quality of education across the system and at all levels. More specifically, there was a need to obtain a national picture representative of all participants in the country and to establish a baseline measure of the quality of education being offered across all schools.

It was within this context that South Africa participated in its first international study; this was in 1994, after the Human Sciences Research Council (HSRC) had been accepted as a member of the IEA. The HSRC is an independent statutory body that conducts research in the human sciences. The HSRC's role in TIMSS was as an independent monitor of the progress South Africa was making after the introduction of new policies in the education system. South Africa did not have the ideal conditions and resources for conducting such a study; for example, the provincial boundaries had not yet been established and there was no central database containing the names of all the schools. The research team had no prior experience in conducting a large-scale assessment nationally. In short, the learning trajectory was considerable and the challenges phenomenal.

The HSRC saw TIMSS as a way to monitor mathematics and science education through a longitudinal study and within an international context. It was an opportunity to compare South Africa's performance in mathematics and science as well as the South African curricula in these subjects with that of other countries and to learn some lessons from the international experience. Finally, it provided an opportunity for South Africa to develop the skills and capabilities necessary to conduct large-scale national assessments of its own in the future.

The HSRC received most of its money through an annual parliamentary grant and so it was well-placed to request funds to conduct TIMSS. Additional concessions from the IEA and later funding from the World Bank eased the financial burden and contributed to funding the project. In this way, TIMSS was conducted in a completely independent manner and was relatively free from political interference in its administration.

### *The Impact of TIMSS Studies*

In South Africa the introduction of the international and regional studies was (and still is) controversial and a sensitive issue in many

circles (Howie & Plomp, 2005b). The results of TIMSS 1995 produced outrage in different circles starting with the Department of Education, which had difficulty accepting the very low performance of the South African learners. The business community and public were shocked that the top-performing learners<sup>6</sup> in the country did not compare with the average learners of the top-performing countries (Howie, 2001). TIMSS 1995 and subsequent international and regional studies revealed not only the damage of past political policies, but also the difficulties of implementing effective change in teaching and learning in South African schools since 1994. South Africa participated in TIMSS 1995 (Howie, 1997; Howie and Hughes, 1998), 1999 (Howie, 2001) and 2003 (Reddy, 2003). However in 2007, it was announced that South Africa would not be joining the TIMSS 2007 study. Controversy surrounded this announcement and a variety of reasons were given by the previous national organizers of the study and the government. The lead research organization suggested that the study was deferred because of the timing of the research, i.e., there was not the time to see results of any policy interventions.

*The South African education system has undergone radical restructuring in its recent past, as several initiatives and interventions have been introduced – each one sharing the common objective of improving teaching and learning in all areas of the curriculum, but especially in mathematics and science. Bearing in mind the strain this intervention has put upon the education system (and more pertinently, the educators themselves), it is recommended that South Africa does not participate in TIMSS 2007, but rather does so in 2011 as this will allow the interventions to become embedded within the education system. This achieved, it would then be more reasonable to measure South African performance in TIMSS 2011 to see how far the country has progressed.* (Reddy, 2003, p.120)

However, the press suggested that the government did not want to

---

6. Learners, pupils and students are used interchangeably in this paper depending on the context.

expose itself to further humiliation. Members of the broader education community regretted this decision as a missed opportunity to continue the external monitoring the quality of education.

South Africa is currently taking part in TIMSS 2011 through its national coordinating centre at the HSRC. The implementation of the study was delayed by a year (data now being collected in August-September 2011) after South Africa placed a national moratorium on international studies until 2011. The data for the southern hemisphere should have been collected in 2010.

### *The Impact of PIRLS 2006 in South Africa*

PIRLS national coordinating centre is based at the University of Pretoria at the Centre for Evaluation and Assessment (CEA); the TIMSS Centre is located in the same city. The Centre for Evaluation and Assessment was established in order to undertake research and offer postgraduate programmes in the areas of systemic evaluation, large-scale assessments, measurement, and assessment at school level. PIRLS 2006 was conducted by the CEA and data were collected in 2006. The co-NRCs, Sarah Howie and Elsie Venter, together with another member of the team, Vanessa Scherman, had all worked on the TIMSS projects at the HSRC, and this experience and expertise proved invaluable for managing PIRLS. PIRLS 2006 faced the challenge of implementing the study in 11 languages at two grade levels (grades 4 and 5). The team feared that the level of the tests would be above the target population of grade 4 and therefore included grade 5 as well. As there are 11 official languages in South Africa, there was a political imperative to test all 11 languages and it was clear that it would not be possible to test fewer languages despite the enormous cost of doing so. The complexity of the study was further underlined with having more than 100 versions of instruments to compile, administer and analyse, involving more than 30,000 pupils in over 400 schools. From the beginning two doctoral students, Surette van Staden and Lisa Zimmerman, worked on the data, one for each grade and

undertook their doctoral studies using the PIRLS data which were completed in 2011 (van Staden, 2011; Zimmerman, 2011). The research team worked very closely with the Department of Education to ensure that the Department was kept abreast of developments as they occurred. Once the preliminary results were available, the Department was briefed on the provisional results.

Despite testing grade 5 learners who were compared to grade 4 learners in other countries, the South African students attained the lowest ranking in the study (302 points compared to the 500 international mean) and were 21 points below Morocco, the only other African country in the study. The grade 4 results were so very poor (about 30 points below the grade 5 results) that the results were not highlighted in the international report's main tables. The very poor results in PIRLS 2006 were lower than could have been predicted and even the 'better' results of grade 5 were a severe blow to the South African government and National Department of Education. The poor PIRLS results released at the end of November 2007 resulted in a national outcry reaching the front pages of every daily newspaper in the country. This placed pressure on the National Department of Education. In Parliament questions were raised about this poor performance and a number of initiatives were set in place. Going into PIRLS 2011, and given the very low results from 2006, the National Coordinating Centre together with the IEA and the International Coordinating Centre for PIRLS decided that South Africa would be better suited to joining the pre-PIRLS study for grade 4 learners in 2011. However, at grade 5 level, children writing in English and Afrikaans would write the PIRLS test, as their 2006 achievement was near to the international mean and the results were considered valid and reliable. PIRLS 2006 now serves as a critical external baseline of reading literacy for grades 4 and 5; against this baseline PIRLS 2011 and future studies will be assessed.

### *Lessons from the IEA Studies and Interventions*

The IEA footprint has been particularly evident in measuring the state



of health of South African education (Howie, 2008). The mirror (one of the purposes of international comparative studies described earlier) presents the country with an 'ugly face' of systematic and systemic failure in education. The problems of the past still haunt the present and the new government has yet to see positive results (in achievement) of the policy initiatives of 1994 and since (Howie & Plomp, 2008). Perhaps the conventional wisdom that it takes 20 years to change an education system is also valid for South Africa. Maybe our expectations were not realistic enough and we are perhaps too impatient. What we should be seeing, however, is the attainment of key indicators that would signal that the basic and essential conditions have been met, so that ultimately achievement will improve.

*Monitoring* our education system through these international studies has revealed that there is a danger that the South African mathematics results as measured in the TIMSS studies are declining. There has in fact been a decline in the scores, although none of the trend analyses have indicated that this is significant – yet. Even more importantly, the monitoring has failed to reveal any improvement despite all the policy changes since 1995; thus, perhaps the conventional wisdom is true: more time is actually required to observe change in our education system. What the monitoring has revealed, however, is that from TIMSS 1995 through to the latest study, PIRLS 2006, there is a clear bimodal distribution of the data (Howie, 2001, Howie & Scherman, 2008): a small cluster of learners at the very top end achieved the highest (international) benchmarks in each study, while the majority (80% and more) did not attain even the lowest international benchmarks. The IEA studies such as TIMSS 2011 and PIRLS 2011 are therefore needed to fulfil this monitoring function.

In South Africa, it has been difficult to discern the extent to which decision-making in education has been based upon or influenced by the findings of the international studies. Whilst there are decisions made and events that follow the international studies and their outcomes, it

is not always easy to categorically link these to the studies themselves. For example, the curriculum revision was underway when the TIMSS 1999 results were released and again after the PIRLS 2006 results were released. Those involved in the curriculum revision process requested the TIMSS 1999 reports and findings and fed these into the decisions made regarding the General Education and Training band curricula for mathematics and science.

The dissemination of the national reports and the subsequent secondary analyses in particular serve the purpose of *enlightenment*. The broader community, politicians, the business community and industry, the media, education organizations, non-government organizations as well as the education system all become involved in the discussion about the results in TIMSS and PIRLS and education in general.

In reaction to questions raised, regarding the quality of literacy and numeracy of primary school learners, in Parliament (National Assembly, Internal Question Paper 9/5/2008) after the PIRLS 2006 results were released, the then Minister of Education (Pandor) mentioned initiatives that were being or had been implemented to address the issue. These initiatives (see Box 1) are viewed as being a direct or indirect impact of PIRLS 2006:

- The Foundation for Learning Campaign, a four-year programme to improve the reading, writing and numeracy skills and abilities of all South African children, was launched in 2008. The Department of Education also focused on providing resources to all schools.
- A Drop All and Read Campaign.
- In addition to the above, the following documents or resources were distributed to schools:
  - The National Reading Strategy document.
  - A Teachers handbook.
  - A Toolkit for teachers.

- An Early Grade Reading Assessment instrument.
- The implementation of the National Policy Framework for Teacher Education and Development.
- To monitor whether learners are improving competency, the Department introduced measures to gather baseline data on literacy and numeracy in the early grades.

### Box 1

In reaction to questions raised in parliament regarding the quality of literacy and numeracy of primary school learners, Minister Pandor had this to say: *“National Systemic evaluations conducted by the Department of Education in 2001 and 2004, revealed low levels of reading abilities across the country. The results of the Progress in International Reading Literacy Study (PIRLS), released in November 2007, found that learners in our schools do not read at the appropriate level in relation to their grades and in terms of their age.*

*Various reasons were provided for this: Lack of access to books in homes, at school and in community, low levels of literacy among the parents and ineffective teaching practices*

*I have responded to these findings through the following initiatives:*

- *On the 18 March 2008 I launched the Foundation for Learning Campaign, a four-year programme to improve the reading, writing and numeracy skills and abilities of all South African children. The Campaign has provided teachers and schools with clear directives on expected levels of learner performance. The focus will be on primary schooling – starting with the Foundation and Intermediate Phases – so that learners acquire and sustain a solid foundation for learning. All primary schools will be expected to increase average learner performance in Literacy/Language and Numeracy/Mathematics to no less than 50% - indicating an improvement of between 15-20% - in the 4 years of the campaign.*
- *The Department has also focused on providing resources to all schools. In the past three years we have provided over 11,000 primary schools with exciting story books, written in all official languages of South Africa, establishing classroom libraries. Through the USAID-funded Ithuba Writing Project, we*

*are distributing 2.3 million locally authored books in the different official languages to schools. All 2.3 million books should be in our schools by the end of the financial year.*

- *Two years ago, I initiated a Drop All and Read Campaign, which welcomed Grade R and 1 learners into education with their own branded bags containing a selection of books that they can read for themselves or that parents and caregivers can read to them. At the heart of the campaign is that in our homes and in our schools children should be able to pick up books that they can read for enjoyment.*
- *We can continue to supply schools with reference materials, which have included bilingual dictionaries. We have also provided all schools with:*
- *The National Reading Strategy document which outlines activities and approaches to promote and develop the reading skills of our learners and*
- *A Teachers handbook entitled Teaching Reading in the Early Grades to assist teachers on methods, approaches and activities to improve their teaching of reading.*
- *A Toolkit for teachers was developed continuing both reading resources as well as guides for teachers; 1000 have been sent to pilot schools countrywide to increase support for the teachers in their teaching of reading.*
- *The Department has developed an Early Grade Reading Assessment instrument, which is currently being used by teachers in selected districts to help us monitor progress in the different schools. The instrument is currently in use for Sepedi, Xitsonga, Tshivenda, isiXhosa and English and during the course of this year the tool will be developed in the remaining five languages.*
- *The implementation of the National Policy Framework for Teacher Education and Development will also address the issue of teacher development and ensure that through the IPET and CPTD programmes teachers are trained to teach effectively.*
- *To monitor whether learners are improving competency, the Department is establishing baseline data on learners' achievement literacy and numeracy in the early grades. As part of the Foundations for Learning, primary school learners will be assessed annually using standardized tests to monitor their progress against the established baseline.*

Source: National Assembly, Internal Question Paper 9/5/2008

PIRLS 2006 also revealed that more than half of the primary schools tested in 2005 had neither school nor classroom libraries, and that more than 50% of children had no access to books at home. In 2008, the Ithuba Writing Project (supported by USAID) distributed 2.3 million books in all 11 languages to the schools. In the first quarter of 2008, Minister Pallo Jordan (whose responsibility included library services) announced that the public library budget would be doubled. The extent to which the PIRLS 2006 had contributed to these initiatives in particular is unknown but they certainly followed the public announcement of the PIRLS 2006 results.

It will be interesting to analyse the PIRLS 2011 data to see to whether there has been improvement since the PIRLS 2006.

### **Conclusion and reflections**

In conclusion, the IEA's footprints can most certainly be found on the African continent, but they are most obvious in South Africa. The main findings of the IEA international studies suggest that the country has not been able to overcome its deprived legacy; that the new policies have not yet been implemented effectively or widely; and that the country has not yet seen the fruits of its more recent initiatives related to education quality. However it does seem that to a large extent its priority of equity of access to education has been achieved, but given the economic and social conditions, continued monitoring of access is crucial.

#### ***Reflections on South Africa's Participation in IEA Studies***

The failure of South African students to meet international benchmarks was revealed in all the international studies in which the country participated. The level at which South Africa's students are able to operate educationally is two to four years behind what the curriculum stipulates. After more than ten years of participating in international assessments of mathematics and science, South African students have

not been able to close the gap on their peers in even the other developing countries participating. The findings revealed the scale of poverty and its impact on children's education when children are only able to eat twice a day, often come to school on empty stomachs and in some cases may go without food altogether. School feeding schemes are clearly a priority now more than ever before. The studies also revealed that a large proportion of South African children are not living with their parents or guardians, but with friends and distant family. This might also explain the lack of interaction with the teachers reported in PIRLS 2006. Although classroom conditions are being addressed--as indicated by the 'School register of needs' project-- they are improving very slowly. Across all the IEA studies reviewed, teachers and principals described the poor conditions and inadequate resources in many schools, the very worst being those in Eastern Cape, KwaZulu-Natal and Limpopo. Large class sizes (average class size found in PIRLS 2006 was 42 at primary level compared to TIMSS 1995 at secondary level where the class sizes were 48 and 49 in mathematics and science classes (Howie, 2002; Howie & Scherman, 2008) continue to challenge the education system at both primary and secondary schools and may well be the reason for the lack of regular classroom assessments being conducted. Average class size continues to be amongst the highest of all the countries participating in the international studies. The costs of reducing these could be considerable, but efforts must be made in this direction if teaching and learning are to improve. Secondary analyses of these international studies (Howie, 2002; van Staden, 2011; Zimmerman, 2011) have revealed the predictors of achievement in South African schools as well as offered deeper insights to the performance levels. One study revealed the importance of language proficiency as the most significant factor predicting mathematics achievement in grade 8 (Howie, 2002, 2003a, 2005a).

The impact of the international studies has differed over the years as has the reaction from various parts of the broader society. The results from the first international study (TIMSS 1995) led to a national con-

cern that there was a real problem with secondary schooling, as the scores were so far below average, but there was no recognition that there might be a problem in primary education. The Department of Education at first adopted an adversarial role as this was a new experience to have the defects (of albeit an inherited system) so publicly revealed by an external assessment. However, as the Department recognized the benefits of such large-scale assessments, their initial rejection converted to an acceptance and collaborative stance. Whilst Prof Asmal (Minister of Education in 2000) declared (at a national launch of the systemic evaluation report) that comparisons, specifically of achievement, were odious, nonetheless apparently note was taken of the TIMSS 1999 finding that there was enormous variation in the number of school days per year (ranging from 120 to 280 days at different schools) and school days as short as four hours. Shortly afterwards, Prof Asmal scheduled 200 school days per year with teachers having to be on the school premises for 7 hours a day. Furthermore, the TIMSS 1999 data on the international and national curricula for mathematics and science was available at the time that the curricula were being revised which means that the committees were able to feed some of the findings directly into the Revised National Curriculum Statement. It was in fact the third Minister of Education after 1994 (Minister Pandor) who appeared to recognize the value of large-scale assessments nationally, regionally and internationally. She, more than any previous minister, supported these assessments whilst recognizing the difficulties related to their implementation and associated publicity. The National Department of Education now manages the SACMEQ studies directly and backed the PIRLS 2006 study, although it did not provide any financial support. South African researchers are increasingly coming together in order to improve the quality of the large-scale studies (e.g., PIRLS 2006 had a national committee comprising leading organizations in the field and experts in reading and assessment), and organizations such as the Joint Education Trust, the Human Sciences Research Council and the University of Pretoria are working more

closely with the National Department of Education in the design, implementation and dissemination of the findings of international, regional and national studies.

Finally, both national coordinating centres for PIRLS 2011 and TIMSS 2011 are working closely with the most recently appointed Minister of Basic Education (who has been supportive of the studies) and also with the government officials in the Quality Assurance and Examinations Unit at the Department. Once again, this collaboration does not involve any financial support and as such the study is an independent research project. The PIRLS 2011 is operating on a skeleton budget and is being financed by the National Research Foundation, Zenex Foundation and SANPAD to a limited extent.

### *Reflections on the African Continent's Participation*

The IEA path is still being forged in Africa. Participation of African nations is growing, and today includes Botswana, Egypt, Ghana and South Africa and Tunisia.

Despite the paucity of African countries partaking in the IEA studies, there is an obvious need for evaluating educational quality on the African continent. The SACMEQ studies emerged in the mid-1990s under the leadership of Neville Posthwaite, Ken Ross and Patrick Griffin, and focused on assessing educational quality in reading and mathematics in countries in southern and eastern Africa. The emergence of the SACMEQ studies could have been seen as competition for the IEA studies; however, the studies focused on the grade 6 level while IEA studies focused on grade 3-4 level. The first SACMEQ study focused on reading, thereafter including mathematics. Interestingly, despite claims that this study would not be comparative in nature, and the first study published each country's results individually, for the second and third studies the results were published together in one report, leading to a comparative study. To date 14 countries have participated in the latest SACMEQ 3 (2005-2009). One may wonder why



these studies were not carried out under the auspices of the IEA and why a new study had to be established, particularly given that Neville was a former Chair of the IEA. It may perhaps be due to the fact that the purposes of the SACMEQ and IEA studies were different.

From the 1990s onwards I have noted and written on issues of cross-national validity and the need for regional studies given the huge difference between Africa and other countries participating in the IEA studies. So why a new set of studies? First, most African countries participating in SACMEQ were not members of the IEA, and so they did not automatically turn to the IEA. Second, there was clearly a need for development in the capabilities to undertake these assessments in Africa and whilst the IEA had run some excellent training programmes, there was no initiative to expand in Africa from this perspective. This may have put a significant strain on IEA resources. Thirdly, the World Bank and UNESCO began to actively encourage developing countries, including those in Africa, to undertake national assessments in their countries (initiated by the Education for All initiative, arising from the Jomtien conference in 1990) and monitor the quality of education offered at primary and secondary schools. Fourthly, under the UNESCO-driven initiative, the focus was on the Ministry of Education taking ownership of monitoring educational quality and there was a perceived need for a more policy-focused set of studies that were targeted more at African countries' policy needs and less on what was possibly perceived as Western-driven studies.

Perhaps in future studies the IEA can consider and reflect on the issues of regional foci and associated studies. What is the future of IEA studies in Africa? Clearly there has been an increase in the number of participating African countries in IEA studies, while simultaneously there has been the emergence of the SACMEQ studies. Some economically stronger African countries (Botswana and South Africa) appear to want to participate in the international studies as well as regional ones and have the best of both worlds. Whilst the Botswanan Ministry of

Education manages both studies from its examination council, South Africa manages its SACMEQ study from the Department of Basic Education as this is a requirement of participating in the SACMEQ study. However the IEA studies are managed from outside the government by a university and a research council (although the latter is largely funded by the government), and therefore are not government driven, which increases the independence of these studies. One may also conclude therefore that the research interests rather than the policy agenda are driving South Africa's participation. For the near future, the participation of African countries in the IEA studies, although limited, is still clearly valued, but funding for future studies is one of the swords hanging overhead.

## References

- Department of Education, (1997). Language in Education Policy. Pretoria: Department of Education:.
- Elley, W. B. (2005). How TIMSS-R contributed to education in eighteen developing countries. *Prospects*, XXXV (2), 199-212.
- Howie S. J. (1997, February). *Mathematics and Science Performance in the Middle School Years in South Africa - a summary report on the performance of SA students in the Third international Mathematics and science Study*. Pretoria: HSRC, Human Sciences Research Council.
- Howie, S. J. (2000, April). *TIMSS-R in South Africa, a developing country perspective*. Paper presented at American Education Research Association Annual Meeting, New Orleans, USA.
- Howie, S. J. (2001) *Mathematics and Science performance in grade 8 in South Africa 1998/1999: TIMSS-R. An executive summary*. Pretoria: HSRC, Human Sciences Research Council.
- Howie, S. J. (2003a). Language and other background factors affecting secondary pupils' performance in Mathematics in South Africa. *African Journal for Mathematics, Science and Technology Education*, 7, 1-20.

- Howie, S. J. (2003b). Conditions in schools in South Africa and the effects on mathematics achievement. *Studies in Educational Evaluation*, 29 (3), 227-241.
- Howie, S. J. (2004). A National Assessment In Mathematics Within An International Comparative Assessment. *Perspectives in Education*, 23 (2), 149-162.
- Howie, S. J. (2005a). System-level evaluation: about language and other background factors affecting mathematics achievement. *Prospects*, XXXV (2), 176-186.
- Howie, S. J. (2005b). Contextual factors on school and classroom level related to pupils' performance in mathematics in South Africa. *Educational Research and Evaluation*, 11 (2), 123-140.
- Howie, S. J. (2010). ICT-supported pedagogical policies and practices in South Africa and Chile: emerging economies and realities. *Journal for Computer Assisted Learning*, 26, 507-522.
- Howie, S. J., & Hughes, C.A., (1998). *Mathematics and science literacy of South African final year students: A summary report of South African students performance in the Third International mathematics and science study*. Pretoria: HSRC, Human Sciences Research Council.
- Howie, S. J., Marsh, T.A., & Hughes, C.A., (2000). Middle school students' performance in mathematics in the Third International Mathematics and Science Study: South African realities. *Studies in Educational Evaluation*, 26 (1), 61-77.
- Howie, S. J., & Pietersen, J. J. (2000). Mathematics Literacy of Final year Students: South African realities. *Studies in Educational Evaluation*, 27 (1), 7-25.
- Howie, S. J., & Plomp, T. (2002). Mathematics Literacy of School leaving pupils in South Africa. *International Journal of Educational Development*. 28 (4), 603-615.
- Howie, S. J., & Plomp, T. (2005a). International comparative studies of education and large scale change. In N. Bascia, A. Cummings, A.

- Datnow, K. Leithwood, & D. Livingstone (Eds.), *International Handbook of Educational Policy*. Dordrecht, Netherlands: Springer, 75-99.
- Howie, S. J., & Plomp, T. (2005b) TIMSS-Mathematics Findings from national and international perspectives: in search of explanations. *Educational Research and Evaluation*, 11(2), 101-106.
- Howie, S.J., & Plomp, T. (Eds.). (2008). Narrowing the Achievement Gap [Special issue]. *Studies of Educational Evaluation*, 34 (2).
- Howie, S.J., & Scherman, V. (2008). Closing the achievement gap between science classrooms and the persistence of historic inequalities. *Studies of Educational Evaluation*, 34 (2), 118-130.
- Howie, S. J., Scherman, V., & Venter, E. (2008). The gap between advantaged and disadvantaged students in science achievement in South African secondary schools. *Educational Research and Evaluation*, 14 (1), 29-46.
- Howie, S. J., Venter, E., & Van Staden, S. (2008). The effect of multilingual policies on performance and progression in reading literacy in South African primary schools. *Education Research and Evaluation*, 14 (6), 551 – 560.
- Howie, S. J., Venter, E. J., van Staden, S., Zimmerman, L., Long, C., Scherman, V., & Archer, E. (2009). South African Children's Reading Achievement: Summary report on PIRLS 2006. Pretoria: Centre for Evaluation and Assessment, University of Pretoria.
- Howie, S. J. & Venter, E. (in press). The impact of PIRLS 2006 in South Africa. In K. Schwippert. (Ed). *The impact of PIRLS around the world*. Hamburg: Waxmann Publishers.
- Moahi, S. (1992). Using evaluation research for policy and practice in Botswana. *Prospects*, XXII (4), 441-447.
- Naumann, J. (2005). TIMSS, PISA, PIRLS and low educational achievement in world society. *Prospects*, XXXV (2), 229-248.
- Plomp, T., Howie, S. J., & McGaw, B. (2003). International Studies of Educational achievements. In D. Stufflebeam & T. Kellaghan (Eds),

- International Handbook on Educational Evaluation*. Dordrecht: Kluwer Academic Publishers, 951-978.
- Reddy, V. (2003). Mathematics and Science achievement at South African schools in TIMSS 2003. Pretoria: Human Sciences Research Council.
- Robitaille, D.F., Beaton, A.E., & Plomp, Tj. (Eds) (2000). *The impact of TIMSS on the teaching and learning of mathematics and science*. Vancouver: Pacific Educational Press.
- South Africa.info. (2006). South Africa: Fast Facts. Retrieved from [http://www.southafrica.info/ess\\_info/sa\\_glance/facts.htm](http://www.southafrica.info/ess_info/sa_glance/facts.htm)
- Van Staden, S. (2011). Reading between the lines: Contributing factors that affect Grade 5 learner reading performance. Unpublished PhD. University of Pretoria.
- Zimmerman, L. (2011). The influence of schooling conditions and teaching practices on curriculum implementation for Grade 4 reading literacy development. Unpublished PhD. University of Pretoria.

## CHAPTER 14

# The Contribution of IEA Research Studies to Australian Education

*John Ainley, John Keeves, Jan Lokan,  
Petra Lietz, Geoff Masters & Sue Thomson*

### Introduction

This chapter is concerned with the contribution of the research studies conducted by the International Association for the Evaluation of Educational Achievement (IEA) to Australian education. During a period of 50 years education across the world has undergone a remarkable transformation. Many of the changes that have occurred have been initiated by the United Nations Organization and its agencies, particularly UNESCO. From its origins within the UNESCO Institute in Hamburg, IEA and its programs have evolved to develop the worldwide conduct of research in education both through the undertaking and reporting of studies as well as the informal training of research workers to participate in these studies. Consequently, the large group of people who have been involved in the IEA studies together with their colleagues have built a new world vision of education in schools. This vision that extends beyond the boundaries of western Europe and North America has been presented in the two editions of the *International Encyclopaedia of Education* with its numerous Handbooks as well as an electronic version named the *Complete Encyclopaedia*.

Space does not permit a more detailed account of the many facets of the two fields associated with the conduct of educational research and the dissemination of scholarly writing about education and the findings of educational research from a worldwide perspective. However, tribute must be paid to the sustained efforts of Torsten Husén and Neville Postlethwaite, both in the leadership of IEA and the editing of the *Encyclopaedia* and *Handbooks*. They and their colleagues have introduced

a global vision of education to a large proportion of the approximately 200 countries that form the United Nations. Moreover, the view taken by IEA in the conduct of research studies has gone beyond decision making and the prediction of the effects of the policies and practices, to the development of an understanding of the processes of education using a modelling approach. This modelling approach emerged from the views of R. A. Fisher, who lived the closing years of his life in South Australia, and Gilbert Peaker, who guided the planning and reporting of IEA research studies during IEA's initial and formative years.

Thus, it is of considerable interest to examine the impact of the IEA research studies of education in one country, namely Australia, over the 50-year period from the early 1960s to the present day. This chapter examines this impact especially in: (a) ideas about education that involve what policy makers, administrators, school principals, teachers, parents and students think about education, (b) practices employed in the classrooms and schools, (c) policies advanced at both the system and school levels of operation, and (d) the conduct of educational research within Australia. The sections that follow in this chapter summarize the contributions of IEA to education in Australia in these four areas, largely through the research activities of the Australian Council for Educational Research.

The account of these contributions is divided into two parts. The first part relates to research activities spanning a period of approximately 25 to 30 years during which IEA was led mainly by Torsten Husén and Neville Postlethwaite, initially from Hamburg and subsequently from Stockholm as well as Hamburg. These initial studies focused, in the main, on the development of an understanding of the forces that influenced educational achievement in nations across the world. In the later period of 20 to 25 years IEA was placed on a much firmer foundation financially and diversified its program of research. This second period was led initially under the guidance of Tjeerd Plomp operating centrally from the Netherlands, with studies administered from many different centres around the world. Thus, in this chapter, it is appropriate

to discuss the influence of these research activities on Australian education in some detail within two separate sections, with many writers contributing to the different sections.

### *The Australian Setting*

The Australian Council for Educational Research (ACER) was invited to participate in the program of research of the International Association for the Evaluation of Educational Achievement (IEA) five years after the first discussions were held in Hamburg in 1958. IEA sought to conduct research into the factors that influenced educational achievement across those countries of the world that had sufficiently strong research skills to enable them to take part effectively in such studies. The invitation came after the First IEA Mathematics Study in 1964 had been planned and less than a year before testing would take place. Since Australia was in the southern hemisphere, testing could be delayed by six months and that gave adequate time for preparation prior to the active involvement of schools. Throughout the period of a little under 50 years the ACER has been the national centre for IEA activities in Australia. It is argued in this chapter that educational policy and practice, as well as educational research in Australia, have profited greatly from the Council's participation in this cross-national program of research.

However, from the outset, in 1963, there was some criticism within Australian educational circles, not only of the tests and other instruments employed in IEA research programs, but also of the whole IEA enterprise (Connell, 1980). Moreover, it was argued that education in Australia was so very different from education in Great Britain, the United States and Europe that comparative research of this kind, which involved an investigation of the outcomes of educational achievement, would have little meaning. Nevertheless, this was an appropriate time for ACER to broaden its interests into the field of comparative education, particularly in the areas of mathematics and science, when there was a major program of reform within Western



countries in the curricula and the teaching of these two subjects. These developments were strongly supported by the mathematics and science teacher organizations across Australia. In addition, in early 1963, the Australian government had decided to change initially to decimal currency and subsequently to metric measures. This was partly because of the impact that computers and calculating devices were likely to have in simplifying computational work, especially in the fields of mathematics and the sciences. This development would involve considerable changes to the school curricula in these subjects. In addition, there was an urgent need to make major changes to the mathematics and science curricula in order to take into account the substantial advances in these fields that had occurred during the first half of the twentieth century.

Education in Australia went through several phases of change in the years that followed on from the early 1960s with the marked expansion of the school and higher education systems. Currently, education in Australia is passing through a further major period of reform. This reform involves the nationwide testing of achievement in literacy and numeracy, the public presentation of the results of performance at the school and system levels, and the provision of greater financial support for schools in need. Furthermore, because of the demand for a more highly skilled work force in Australia, as well as the increasing mobility of the population between the States and Territories of Australia, a national curriculum is currently being developed for introduction in successive stages across the country. The first phase covers the areas of the curriculum in English, mathematics, the sciences and history and appropriate testing programs are likely to be introduced to monitor student learning in these subject areas. This work is under the guidance of Barry McGaw, who was formerly a Director of ACER, and the Australian representative at IEA General Assemblies. He is also an editor of the third edition of the *International Encyclopaedia of Education*.

## **The Contribution of IEA to Australian Education before 1990**

The ideas presented and discussed in this first section draw heavily on the ideas advanced by Torsten Husén, Neville Postlethwaite, Gilbert Peaker and their many colleagues within the IEA group who contributed to IEA publications, the *International Encyclopaedia of Education* and less formal writings on the planning, conduct and findings of IEA research studies.

### ***The Contribution to Ideas about Education***

During the first 30 years of its operation, from 1930 to 1960, the work of the Australian Council for Educational Research focused on the educational and psychological assessment of students, and catering for individual differences, as well as the advancement of new and progressive educational ideas. Arising from the evaluation program of the Eight Year Study in the United States, Ralph Tyler summarized the ideas underlying this study in his *Basic Principles of Curriculum and Study* (1949). This was followed by the *Taxonomy of Educational Objectives* compiled by Benjamin (Ben) Bloom and his colleagues in 1956 for the cognitive domain (Bloom, 1956), and after its successful reception, for the affective domain by Krathwohl, Bloom and Masia, (1964). Subsequently, in 1971, Ben Bloom published the *Handbook on Formative and Summative Evaluation of Student Learning* (Bloom, Hastings & Madaus, 1971), and IEA ran a training workshop at Gränna in Sweden, based on Ralph Tyler's monograph and the Handbook. Ben Bloom visited Australia and the ACER in the latter part of 1972. The ACER responded to these new ideas about evaluation through conducting a postgraduate course in the School of Education at the University of Melbourne led by William C. (Bill) Radford and S. S. Dunn for the training of students and teachers in the conduct of evaluation in schools. Over time the orientation of the research program carried out by ACER and the perspective of much

educational research conducted within Australia has moved away from assessment and concern for individual differences. Its research program has become oriented towards a focus on schools and the evaluation of the learning experiences provided by schools, and has been greatly influenced by the evaluation program developed by IEA.

The tasks of evaluation employed in the IEA studies may be considered to have six theoretical components: (a) curriculum objectives and curriculum design, (b) a model of school learning, (c) curriculum implementation at three levels (namely systemic, school or classroom, and individual levels) leading to a multilevel structure of operation, (d) modelling and modifying educative processes, (e) the multi-disciplinary nature of educative processes, and (f) the investigation of change over time at the three different levels. In the sections that follow, brief consideration is given to these six components and the changes in theoretical perspectives in Australia that have arisen from the IEA studies. These six theoretical components are considered in greater detail in two other chapters in the two volumes that mark the first 50 years of IEA's program of research.

### **Curriculum objectives and curriculum design**

No longer was the curriculum of the schools and the school system viewed only in terms of knowledge in specific content areas. Within the Tyler rationale promulgated by Ben Bloom, cognitive skills and affective orientations were developed that interacted with the content aspects which increased in importance across the levels of schooling. Learning was not the mastery of specific criterion-referenced tasks, but an ongoing process throughout schooling and life with identifiable standards at recognized stages. This required a cross-sectional mapping on a content-process grid of the curriculum objectives. Thus, the design of a subject curriculum involved the sequencing of the cross-sectional maps across the stages of learning, together with identified standards at each stage.

### **A model of school learning**

The model of school learning advanced by John B. Carroll has guided many IEA research studies. Carroll (1963) proposed a model of school learning that involved five factors which were grouped under two headings: (a) factors involving time needed for learning, and (b) factors involving time spent on learning. Factors that involved the individual were (i) **aptitude** - the time needed to learn the task under optimal conditions of instruction, (ii) **ability** to understand instruction, and (iii) **perseverance** - the time the student was willing to engage actively in learning. Factors that involved external or classroom conditions were: (iv) **opportunity to learn** - the time allowed for learning, and (v) **quality of instruction**. The quality of instruction also involved the sequencing of tasks prescribed for learning across the stages of student development and growth. IEA research studies have supported strongly the importance of Carroll's ideas, and where once they were challenged, they would now seem to be well accepted in the planning of curricula within schools.

### **Curriculum implementation at three levels**

From the Gränna workshop conducted by IEA in 1971 strong support was provided for identifying at three separate levels: (a) the **prescribed curriculum**, (b) the **opportunity to learn** or the **implemented curriculum**, and (c) the **achieved curriculum** (Keeves, 1974). Each of the three levels was linked to the school system, the school or classroom, and the student, respectively, and also with the conditions of learning in the country or system, the school or classroom and the home. These views of the curriculum were clearly identified at the three separate levels that pervaded all school learning and teaching. As a consequence it was no longer meaningful to ignore the multilevel nature of educational situations, not only in research studies, but also in the making of policy as well as in practice.

### **Modelling and modifying educative processes**

Educative processes were seen to be complex and took place in natural settings. In general, while an intervention was frequently possible, an experimental study would destroy the realism required for the meaningful investigation of the processes of education, and subsequent advancement of teaching and learning. Certain factors operated through or were mediated by other factors. In addition, certain factors interacted with or were moderated by other factors, so that, for example, some students within a classroom group might be influenced differently from other students within the group when classroom conditions were modified or changed. These complex multivariate and multilevel operations could not be ignored and needed to be examined both in research studies and in practice.

### **The multidisciplinary nature of educative processes**

During the first half of the twentieth century it was widely accepted that educative processes were governed by psychologically based operations. The involvement of psychology could not be denied, but the processes of education were more complex. The educative processes drew on many disciplines for a coherent advancement of education. Sociology, economics, politics, history, demography, philosophy and the sciences all contributed. Education thus became a multi-disciplinary field of inquiry. Moreover, it contributed in an active way to all other disciplinary and inter-disciplinary studies. These ideas emerged from the discussions at the Lake Mohonk Conference (Super, 1967) that are presented in two other chapters of the two volumes on the history of IEA.

### **The investigation of change over time at the three different levels**

The recognition of several levels of operation for the processes of education facilitated the investigation of change and development. Education was not a static operation but a dynamic one. Consequently,

any investigation of situations in the field of education needed to examine change and development. While the evaluation studies conducted by IEA were initially cross-sectional and static in nature, the continuing conduct of studies over a 50-year period led to a new approach that involved the study of change not only with respect to learning and teaching, but also with respect to planning for the future and the reviewing of the past.

### *Impact on Practice at the School and Classroom Levels in Australia*

In this section the contributions of IEA studies to the practice of education at the school and classroom levels in Australia are considered. Five areas are identified where IEA studies have had a recognizable impact on educational practice in Australia by providing evidence of significant effects, namely: (a) student attitudes and values, (b) student views of the teaching and learning processes, (c) time and opportunity to learn the implemented curriculum, (d) time spent on homework, and (e) class size with respect to individualized or whole group instruction.

The teaching practices that benefited student learning were difficult to identify in the early IEA studies. This was largely because of the difficulties encountered in measuring change in achievement over time, and in estimating the magnitudes of the effects operating separately at the student, classroom and school levels. IEA scholars became involved in the development of procedures that had the capacity to overcome these analytical problems and the ongoing nature of the IEA program of research would seem to indicate that future studies should lead to substantial progress.

#### **Students' attitudes and values**

The First IEA Mathematics Study in 1964 sought to assess students' attitudes and values and to examine their relationships with achievement on the mathematics tests. Three scales were constructed to assess attitudes towards: (a) mathematics as a process, (b) difficulties of

learning mathematics and (c) the place of mathematics in society. In addition, two more general attitude scales were constructed towards: (d) school and school learning, and (e) man and his environment. The scales employed 'agree' and 'disagree' responses and the scales were examined as Guttman scales with respect to: (a) the constructors' judgements of the ranked order of the items, (b) the variation in ranked order of the items across countries, and (c) problems encountered in the translation of items to languages other than English. Items exhibiting inconsistencies were eliminated and coefficients of reproducibility were estimated for each country. In addition, a simple scale was constructed to assess interest in mathematics. Subsequent IEA studies have always included attitude and value scales that were appropriate to the subject areas under survey, although in general Likert scaling procedures have been employed. The difficulties encountered in analysing the data obtained from the attitude and value scales arose from the problems involved in modelling the relationships between the attitudes and achievement outcomes at student, classroom and school levels. In addition there were problems associated with differences between countries arising from the translation of attitudinal statements from English into other languages. Sex or gender differences in effect size that were statistically significant were generally recorded, but the effects of such differences on achievement were not found to be consistent across countries.

Problems were also frequently found to emerge from the failure of students to answer all items in the attitude scale questionnaires, arising from the need to ask a sizeable number of items for each scale. Rasch scaling procedures were then developed in Chicago by Andrich (1997) and Masters (1997), as graduate students, that also provided for scaling with omitted item responses, since the scale values estimated were independent of the individual items employed. Nevertheless, while there was little doubt that attitudes had an important role in their influence on educational outcomes, these effects were also likely to moderate or have cross-level interaction effects on changes in achieve-

ment over time. Consequently, time series or longitudinal studies were required to obtain stronger findings for the effects of attitudes on achievement. With cross-sectional survey data, it also became possible to use two-stage least squares procedures and maximum likelihood procedures to estimate the sizes of effects for the possible reciprocal relationships between attitudes and achievement. Moreover, with multilevel analysis procedures it became possible to examine effects not only at the student level but also at the classroom and school levels. The investigation of the effects of attitudes and values on teaching and learning in education is now taking place in Australia using procedures advanced by Andrich and Masters that involve Rasch scaling.

### **Student views of teaching and learning processes**

In the Mathematics Study two descriptive or view scales were employed to assess student views of: (a) mathematics teaching, and (b) school and school learning. These scales, like the attitude scales, were initially treated as Guttman scales. Likewise, in the First and Second IEA Science Studies it was decided to assess the emphasis in science teaching on inquiry through the use of scales concerned with: (a) scientific investigation, and (b) involvement in science practical work. In the Science Studies these two scales were processed as Likert type scales. It would now be possible to obtain scores for these scales using Rasch scaling procedures. The scales were trichotomous with 'often' scored 1, 'sometimes' scored 0.5, and 'never' scored 0. The Scientific Investigation scale was answered by teachers, and the Student Practical Work scale answered by students. Considerable work into the use of descriptive scales has been undertaken at the Science and Mathematics Centre at Curtin University in Western Australia by Fraser and his colleagues (Fraser, 1997).

### **Time to learn and opportunity to learn the implemented curriculum**

In the First IEA Mathematics Study in 1964 and the First and Second



IEA Science Studies in 1970-71 and 1983-84, respectively, it was considered important to examine the extent to which the prescribed curriculum had been implemented in schools through assessing whether the opportunity to learn the items included in the tests had been given to the students. Thus, the science teachers of the students tested were asked to assess the extent to which their students had had the opportunity to learn the content of each of the items included in the tests. Consequently, it was possible not only to examine the differences between countries at the country or system level, but also to examine the effects of opportunity to learn on the average level of achievement of each school. Such relationships were more likely to be found in countries where the schools had greater freedom to develop their own science curricula. Where there was a centrally prescribed curriculum for all schools in a country it was less likely that a strong relationship would be found when other factors were taken into consideration. At the 10-year-old level, small but significant relationships were found for Australia and England. Likewise, at the 14-year-old level in Australia, England and Italy, relationships were found to be associated with the time spent on the study of science. These were the countries where schools had greater freedom to decide what was taught to students in science courses. However, it was between countries, rather than between schools within countries, that opportunity to learn had more substantial relationships with student achievement (Keeves, 1992a, 1992b, 1995).

### **Time spent on homework**

The evidence from the First IEA Mathematics Study at the lower secondary school level, but not at the pre-university level, showed that time spent on homework in all subjects per week was positively correlated with mathematics achievement with a median correlation of 0.22, and all correlations recorded were positive (Husén, 1967; Keeves, 1995). However, no consistent patterns of results were reported for the relationships between time spent on mathematics homework and mathematics achievement. The analyses conducted in the First IEA

Science Study at both the between school and the between student levels established the importance of time spent on homework in nearly all countries at the lower secondary school level. While the time spent on homework in all subjects had positive simple correlations with science achievement, inconsistent patterns of results were recorded at both school and student levels of analysis for time spent on science homework. Similarly, results recorded from the Second IEA Science Study confirmed that the length of time spent on homework in all subjects, and time spent on science homework had, in the main, a positive relationship with achievement in science (Postlethwaite & Wiley, 1992). Whether these effects arose from the extended use of instructional time, or whether they were a consequence of a student's academic motivation and perseverance, remained unclear. The use of multilevel analysis and time series or longitudinal data could be expected to yield stronger results, as would the examination of the data using multivariate regression and path analysis models.

### **Class size and whole group instruction**

The First IEA Mathematics and the Science Studies in Australia were generally found to yield positive correlations and regression coefficients between class sizes and achievement (Husén, 1967; Keeves, 1992a). Nevertheless, the results from studies conducted within Australia using multilevel analysis procedures involving least squares regression, and also HLM and MPlus, have shown positive relationships between class size and achievement that might indicate the possible use of practices involved in the selection within schools of students into large and small classes on the basis of ability (e.g., Larkin & Keeves, 1984).

The IEA Classroom Environment Study in Australia conducted in the 1980s by Bourke (1984) provided an opportunity to examine the effects of class size on the teaching and learning of primary school mathematics. This study found that while class size was negatively related to student attitudes towards the importance of mathematics it was positively related to student achievement at the student level. However, in

the examination of the data at the classroom level, when student ability was included in the regression model, a suppressor relationship was recorded and the sign of the regression coefficient ( $b=-0.24$ ) was associated with an initial positive correlation coefficient ( $r=+0.18$ ). Such a relationship was considered difficult to interpret, in part because multilevel analysis procedures were not available at that time. Both Fordham (1983) and Bourke (1984) provided evidence from classroom observation in this study that many teachers, when working in smaller classes, did not shift from group methods of instruction to more individual methods of student learning. Bourke argued, "An explanation probably revolves around teacher as well as student selection for large and small classes, and a possible lack of knowledge by teachers of instructional practices that could be used to take advantage of smaller classes" (1984, p. 215).

Lindsey (1974), working in France, undertook an examination of relationships with data from the First IEA Mathematics Study using response surface analysis as a special case of multiple regression with the data for 13-year-old students. He reported that the highest mean scores for the English and Scottish samples were found for class sizes in the range of 20-30 students. This type of analysis does not appear to have been attempted with Australian data, but Lindsey's analysis illustrated the modelling strategies that would be carried out in order to examine particular issues with data from natural settings.

### ***Impact on Policies Advanced at the System and School Levels of Operation***

Australia is a nation that is formed by the federation of six States and two Territories, and while the State and Territory Governments have responsibility for the provision of education in schools, the Australian Government provides substantial financial support for education at all levels. The policy-related issues raised by IEA in its cross-national studies also apply to the making of educational policy at both the state and national levels across Australia.

### **Retentivity and attainment**

In 1964, the First IEA Mathematics Study in its National Case Study Questionnaire sought information on retention rates across age and grade levels for both the state school systems and the country as a whole. This work revealed gaps in the information available as well as striking results, because the immediate post-war cohort of children were, in 1963 and 1964, reaching the upper-secondary school level. A monograph, titled *Staying Longer at School*, that was prepared by the Director of ACER and IEA Council member Bill Radford (1966), showed the changes that were occurring within Australia. This information, following improvement of the procedures employed for the collection of data, provided the standard indicator of the outcome of educational attainment in Australia. The index was known as the 'retention rate', and was calculated at the age and grade levels of secondary education.

In the subsequent reports of both the IEA Mathematics and Science Studies for Australia it was possible to examine the effects of the differences in retention rates between regions and school systems, as well as between males and females, on achievement in mathematics and in the separate fields of science, namely, biology, chemistry and physics. Moreover, following the work undertaken by Postlethwaite (1967) into an index of 'yield' of school systems at the terminal secondary school stage, estimates of yield were examined for the Australian States over time between the successive studies in mathematics and science. These findings became of considerable importance in attempts made nationally to raise the levels of qualified personnel for working in scientific and technologically based occupations (Jones, 1988). In addition, in the First Australian Mathematics Study report, Keeves (1968) showed that the relationship between estimates of retentivity and change in levels of achievement between the middle secondary school and terminal secondary stages was linear and inversely related, and was consistent with the retentivity model advanced in the First IEA Mathematics Study report by Walker (1967). This work has proved to be highly controver-

sial when extended beyond the State school systems to those systems associated with school type, between the Government, Independent and Catholic Schools, as well as between the States and Territories, to the extent that information that would allow comparisons between school types to be made and linked to achievement outcomes was not released until changes in reporting practices were introduced in 2010.

### **Socioeconomic status differences and issues of equity**

In the First IEA Mathematics Study information relating to socioeconomic status was collected on a nine-point occupational scale. Relationships were examined between scale category levels and rank scaled score values and student achievement, interest, and attainment from the Grade 8 to the terminal secondary school levels (Keeves, 1968; Keeves & Radford, 1969). Such results had not been estimated previously in Australia with national samples of students. An Australian occupational scale, based on the occupational level of skill and skill type was not released until 1965, well after the data collection for the IEA study was completed. Significant relationships (Keeves & Radford, 1969) were recorded for mathematics achievement, interest in mathematics, and attainment to the terminal secondary school stage indicating the effects of social class and status of the home on student performance at school. While no significant difference was reported for mathematics achievement between students living in metropolitan and rural areas in Australia, differences in attainment at the terminal secondary school stage were found that indicated the disadvantage experienced by students living in rural areas (Rosier, 1978). This work suggested the operation of significant effects of the social class composition of the community served by a school. In recognition of this influence on the outcomes of educational achievement and attainment, the Interim Committee for the Australian Schools Commission introduced compensatory programs at the school level in order to provide for social disadvantage and lack of equity. This policy was later examined in detail by Ken Ross (1983), who developed several indices that were formed using

census data as a basis for funding schools to overcome the effects of social disadvantage and to raise the quality of schooling in Australia.

### **Sex differences and gender effects**

The First IEA Mathematics Study in 1964 (Moss, 1982) and the First IEA Science Study in 1970-71 (Keeves, 1973; Keeves & Read 1974) reported clear evidence of sex differences in achievement, attitudes and participation in the learning of mathematics and the sciences across countries. In mathematics, the sizes of effect associated with gender differed between countries in a manner that appeared to be related not to genetic factors, as had sometimes been assumed, but to teaching and curricular factors and possibly to the roles of women in society. This societally based explanation was subsequently confirmed cross-nationally by Baker and Jones (1993). The effects recorded in the Mathematics and Science Studies, with, in the main, boys performing better than girls in mathematics, and with greater sex differences reported in favour of boys in physics than in chemistry and biology, would appear to be widespread across the countries involved in the IEA studies. These findings were strong enough to require compensatory programs. It was recognized that such programs could be implemented within schools to increase the participation and performance of girls in the areas of both mathematics and the sciences. The data collected in the Second IEA Science Study showed significant reductions in the magnitudes of the sizes of the sex-difference effect in Science (Keeves, 1992a). Likewise, between the first and second studies there were decreases in the gender ratios for yields in mathematics and science recorded at the terminal secondary school stage (Moss, 1982). The evidence that changes were occurring over time not only in Australia, but also in other countries, indicated, very convincingly, that these differences were to a significant extent societally based. The introduction of educational policies and programs to overcome what was termed 'gender bias' would appear to have had recognizable effects. Nevertheless, the situation could be more complex than might have

been implied by the apparent effects of compensatory programs insofar as major changes were also taking place at a societal level in many countries of the world. Moreover, with the emergence of a worldwide shortage of trained personnel to work in scientifically and technologically based fields, any reduction in the attraction of these fields for any students, which could occur for boys from 'feminized' curricula, might have serious consequences.

### **The influences of curricular time and participation rates**

In Carroll's (1963) model of school learning referred to in an earlier section in this chapter, the five factors advanced in the model could also be grouped under two headings: (a) factors influencing time needed for learning, and (b) factors influencing time spent on learning. Where once the relationship between curricular time and achievement in school learning was questioned, it later came to be widely recognized. A significant relationship was shown (Keeves, 1968) across the Australian States from the First IEA Mathematics Study data between achievement and two variables involving time, namely, total prescribed curricular time spent in learning mathematics during the first seven years of schooling, and prescribed time spent in learning mathematics during the current and eighth year. The findings from the Second IEA Science Study in 1983-84 confirmed the operation of a positive relationship between curricular time and achievement (Keeves, 1992a, b). However, at the 10-year-old level for the seven industrialised countries, including Australia, under survey at this level, the straight line graph did not pass through the origin, and it was argued that scientific knowledge was also probably gained from reading books, magazines and encyclopaedias and from the mass media, particularly from television.

A further factor of some importance arose where the learning of a subject such as a specialist science, or an alternative type of mathematics as occurred at the upper levels of secondary schooling and where the study of the subject under survey was optional. In such situations it became necessary to distinguish between the retention rates, which

involved the proportion of an age or grade cohort remaining at school, and the participation rates, which involved the proportion of an age or grade cohort participating in a subject or an alternative type of subject at the level under survey. Failure to relate the index to an age or grade cohort by merely considering the proportion of those students enrolled at school at the level under survey confounded the examination of such relationships with achievement. Consequently, relationships between retention rates, as well as participation rates, and achievement needed to be estimated. Furthermore, the examination of such situations for the significance of linear relationships gave rise to outliers that had to be discarded. Generally, in the examination of relationships both between the Australian States and between countries, significant relationships were recorded. Clearly 'dropping out' from school and 'dropping out' from a subject or studying an alternative type of subject involved no time given to the study of the subject or a reduced proportion of curricular time given to the study of the subject. Allowance needed to be made for such effects if meaningful results were to be obtained. Where such allowances were made, meaningful results were generally found but with outliers indicating possible anomalies in the information provided (Keeves, 1992a).

### **Curricular differences among the Australian states**

In the subject fields of mathematics and the sciences, detailed analyses of the associated curricula were required for IEA studies prior to participation in the testing program. Analyses were undertaken (and published in the Australian national reports) that involved the completion of a content-process grid, together with a rating on a scale with three levels of importance given for each particular cell of the grid (Keeves, 1968; Rosier, 1980). In general, the findings from these analyses showed the uniformity of the mathematics and science curricula across the Australian States and Territories so that a national view of the curriculum for Australia in these fields was meaningful. However, the different ages and conditions of entry to schooling in the different



States and Territories, as well as the different structures of the school systems of the States and Territories, served to confound the comparisons made at specific grade levels. Consequently, sound comparisons related to levels of performance in the different States and Territories have been made through the employment of age samples rather than grade samples. Where this procedure was not adhered to, highly confusing and misleading results were frequently obtained. This led to a reluctance to analyse and present the findings of IEA studies with respect to specific States and Territories, except in particular circumstances where the States and Territories could not be identified.

Relationships have been explored between curriculum content ratings and student achievement, with consistent results which indicated that, where differences in achievement were recorded, both in mathematics and the sciences, they were related to differences in the curricula. However, not only were the curricular differences among the States and Territories small, but the differences in mean levels of achievement and associated variances were, in general, also small (Rosier, 1980; Rosier & Banks, 1990; Rosier & Long, 1991). Thus, the lack of uniformity among the States and Territories in curriculum specification served to accentuate superficial differences as far as teachers and school principals were concerned. Nevertheless, since students left school at different ages and at different grade levels, the variation in student performance across Australia was larger than might have been expected. These issues warranted examination and subsequently have been given consideration in the development of the national curricula, as well as the development of greater uniformity in policies for the States and Territories of Australia.

### *Contributions to the Conduct of Research and Evaluation in Australia*

The IEA studies have been concerned with research in the fields of education through the evaluation of educational achievement. The view of 'evaluation' was firmly grounded in the rationale developed

by Ralph Tyler and supported by the work of Ben Bloom and Richard (Dick) Wolf. The methods of survey research flowed from the statistical ideas of R. A. Fisher that were strongly endorsed by Gilbert Peaker. Seven methodological and analytical procedures have been employed by IEA in the period during which the Association has been conducting cross-national evaluation studies into the problems relating to teaching and learning in the fields of education in schools. These procedures have greatly influenced the conduct of educational research within the ACER and within Australia.

### **Curriculum objectives and their analysis**

As already noted, the starting point was the identification of the curriculum objectives of the school systems and their schools. The Taxonomies (Bloom, 1956; Krathwohl, Bloom, & Masia, 1964) prepared by Ben Bloom provided the foundations for this work, and they were discussed in the Handbook (Bloom, Hastings & Madaus, 1971) that formed the basis for the Gränna Workshop conducted in Sweden by IEA. In the field of science, in particular, the curriculum analysis procedures presented in the Handbook were the basis for the two IEA Studies of Science Education in 1970-71 and 1983-84 as well as the studies of Mathematics Education in Australia, as presented in detail in the reports of these studies both at an international level (Comber & Keeves, 1973; Husén, 1967; IEA, 1988; Keeves, 1992a; Postlethwaite & Wiley, 1992; Rosier & Keeves, 1991), and at the national level (Rosier, 1980; Rosier & Banks, 1990; Rosier & Long, 1991). They supplied detailed information for the examination of changes in the teaching of science and mathematics over a 14-year period. Moreover, both the national science reports, as well as many of the international reports, were provided by the IEA team of scholars working from within Australia (Keeves, 1992a, 1992b, 1995; Rosier & Keeves, 1991).

### **Multiple choice testing and the use of attitude and view scales**

The problem confronting those involved in the development of instru-

ments for the evaluation of educational achievement was to assess the performance of very large numbers of students. While multiple choice tests could be readily processed through machine scoring procedures, it was recognized from the outset in the First IEA Mathematics Study in 1964 (Husén, 1967) that constructed response items were also required in addition to multiple choice items. In subsequent studies greater emphasis was placed on so-called 'authentic' assessment and performance procedures, such as the assessment of practical work in science (Doran & Tamir, 1972). Not only was the assessment of practical work a new procedure in science education in Australia, but the use of multiple choice tests and machine scoring of answer sheets were completely new procedures in the assessment of mathematics in Australian schools in 1964, as they were in many other countries that participated in the First IEA Mathematics Study.

### **Sample design, sampling errors and practical significance**

Gilbert Peaker's extensive experience in the design of samples and in the estimation of errors of sampling and assessment in England shaped this aspect of IEA research studies for many decades (Husén, 1967; Peaker, 1975). The procedures of stratification and the allowance for the effects of stratification in the analysis of data were completely new in Australian education. In addition, the use of cluster sample designs and two stages of random sampling with probability proportional to size sampling of schools and limited numbers of students drawn randomly from within each school were also new. Moreover, the problems associated with the estimation of standard errors that involved allowance for the sample design were largely unknown in educational research in most IEA countries. Furthermore, the use of the idea of the size of an effect and practical significance became a necessary procedure for the examination and reporting of results across many countries where the samples differed greatly in size and design. Procedures for the weighting of data were better known, but not widely employed until recently.

### **Indicators of socio-educational advantage and disadvantage**

It was a bold move in the First IEA Mathematics Study to include the use of a nine-point occupational scale, and the use of this scale in the reporting of results was undertaken with serious reservations in the comparisons across countries. However, from the work of Gilbert Peaker (Peaker, 1967) in the Plowden Study, the effects of socio-educational disadvantage were acknowledged. Such effects were widely debated in the United States following the release of the *Equality of Educational Opportunity Study* by Coleman in 1966. In the IEA Six Subject Study the indicators of social advantage and disadvantage were extended beyond ideas of class and occupation to include such aspects as family size, number of books in the home and daily receipt of a newspaper, with separate indexes formed for each country (Peaker, 1967). IEA studies in Australia were at the forefront of such work, although recent indicators in use in Australia have moved away from an educational and occupational orientation to an economic one. Moreover, there has been an increased emphasis on the characteristics of the community served by a school, following the work in Australia by Ken Ross (1983) with indicators developed at the school level, rather than the consideration of effects operating at the individual student level.

### **Longitudinal and trend studies**

The early studies conducted by IEA were essentially cross-sectional in nature. However, as time passed it became possible to introduce different forms of longitudinal studies. Trend studies (Keeves, 1992a, b), which tested the same age or grade group using different samples at the same time points and at different points in time, were undertaken with Australian data that built on studies that were primarily cross-sectional (Comber & Keeves, 1973; Keeves, 1968; Rosier, 1980). Time series studies that followed the same sample across successive time points were also undertaken with two time points (Keeves, 1972) within an Australian situation that changed over time. However, no stud-

ies have been conducted by IEA within Australia with a deliberately planned and experimental intervention.

### **The examination of growth and learning trajectories**

A strong feature of hierarchical linear modelling (HLM) developed at the IEA centre in Chicago (Bryk & Raudenbush, 1992) is its capacity to handle growth and learning trajectories at the micro-level and to examine cross-level interaction effects operating at the student level or meso-level, and interaction effects from the schools at the macro-level. These cross-level interaction effects have been referred to as moderating effects and they have added an important new dimension to the analysis of data in IEA linked studies using Australian data collected in the 1960s (Darmawan & Keeves, 2006; Larkin & Keeves, 1984).

### ***Rasch Scaling and the Measurement of Educational Outcomes and Explanatory Variables***

The possibility of undertaking longitudinal, time series and trend studies in the IEA program of research required that the data for criteria, as well as explanatory variables that were replicated over time, needed to be equated not only over time by horizontal equating, but also over grade levels of schooling by vertical equating. Moreover, the scales employed needed to be robust and interval in nature and constructed to be independent of the items used and the students employed to calibrate each scale. At the initial planning meeting in Hamburg in 1958, Georg Rasch was present and he had subsequently travelled to the United States to visit the Department of Education at the University of Chicago in 1960 to inform that centre about what came to be known as Rasch scaling. Ben Bloom and his colleague Ben Wright accepted Rasch's ideas with enthusiasm and this new approach to measurement in education became a core component of the courses in Measurement, Evaluation and Statistical Analysis conducted for graduate students in the University of Chicago. Students came from many parts of the world to take these courses, including many from

Australia. Within Australia the Australian Council for Educational Research introduced this development in educational measurement into the graduate courses conducted at the University of Melbourne.

In addition, to mark the fiftieth year since its foundation the ACER organized a conference on 'The Improvement of Measurement in Education and Psychology' (Spearritt, 1982). Bob Thorndike and Bruce Choppin, who were heavily involved in IEA studies in the United States and England, respectively, gave keynote papers. In addition, in anticipation of the use of the Rasch latent trait measurement model in the equating of science tests in the Second IEA Science Study and in tests of literacy and numeracy in the ACER's current program of research, George Morgan (1982) presented a paper on the equating of scholastic aptitude tests over time. The interest in Australia in this new approach to measurement in education was strong. Andrlich from Western Australia gave a paper and Ray Adams, Barry McGaw and Mark Wilson, who attended the conference, have continued on to make their mark in this field of research in education. Subsequently, Geoff Masters returned to Australia from studying in Chicago to become heavily involved in the technical side of the IEA research program. Australia currently has centres in Adelaide, Brisbane, Melbourne, Perth and Sydney where research is conducted into the problems of measurement in education.

### **ACER's Involvement in IEA Research Studies since 1990**

During the first 30 years of operation both overseas and in Australia, IEA scholars laid the foundations for an ongoing program of research and evaluation. Since the early 1990s, this work and the number of national centres involved in the IEA program of research has grown markedly, together with the number and nature of the fields of inquiry and the types of studies. The ACER has extended its interest and participation into four clearly identified fields, each with its own methods of investigation. In the sections that follow, brief accounts of these

fields of inquiry are given, together with the consequences of their findings for both policy and practice, as well as the conduct of educational research in Australia. These four fields are listed below:

- (a) Trend Studies in the Fields of Mathematics and the Sciences;
- (b) Probing Mathematics and Science Teaching Practices More Deeply;
- (c) Information and Communication Technology in Australian Schools;
- (d) Development of Measurement in Educational Research: Rasch Scaling.

### *The Conduct of the IEA Research Program in Australia*

As the major player in almost all of the IEA studies carried out in Australia, the ACER has used management and dissemination structures that maximise the studies' potential to influence curriculum and practice in Australian education. Each study has had a Steering Committee with one or more representatives from each of the six States and two Territories that to date have had responsibility for school education within their jurisdiction, as well as representatives from the Commonwealth Department of Education (variously named over time), which has had responsibility for certain specific programs. Typically, educational bureaucracies, universities, subject-matter associations and teachers' unions have provided the Steering Committee members. Thus, educators with expertise and influence in curriculum initiatives have gained awareness of the scope and aims of IEA studies and helped to guide the local implementation of these studies. They have also been involved in the interpretation of results and in the discussions of their implications for Australian education.

Full reports and summaries of the results of the IEA studies have always been produced, but, to reach wider audiences, the ACER has greatly increased the range and nature of its dissemination activities

over the past 20 years. The ACER Research Conferences series began in 1997 with a two-day residential conference in Melbourne on *Raising Australian Standards in Mathematics and Science: Insights from TIMSS* (ACER, 1997). Papers covered the study itself and a summary of results. Implications for mathematics and science education in Australia were raised by key mathematics and science educators from Melbourne and Monash Universities. A paper on mathematics and science education in Singapore was presented by a senior curriculum officer from their Ministry of Education. Furthermore, a keynote address on mathematics education in three countries, with results from the first use of video cameras as a pilot IEA study within TIMSS, was presented by James Stigler, of the University of California in Los Angeles. The conference, which was well-attended by administrators, academics and some principals and senior teachers from across Australia, concluded with a thoughtful analysis of how the TIMSS results could be used to change, reinforce or extend current practice in mathematics and science education. This paper, entitled *The way forward: Suggested responses for Australian education systems*, was presented by Rice, of the New South Wales Department of Education.

ACER always prepares Press Releases when an IEA study report is about to be made public, and these usually result in newspaper articles, radio interviews and television interviews. While the media's main interest usually focuses on Australia's rank order position on the achievement ladder, the managers and researchers interviewed take the opportunity to emphasise the important implications of the results for the Australian education systems and their programs. Since the inaugural ACER Research Conference in 1997, there has been a similar conference each year on a topical educational issue; one or more papers arising from an IEA study is always featured where the study is relevant to the conference theme. In addition, presentations have been given by ACER project staff at a wide range of conferences, both domestic and overseas, including the American Educational Research Association, the National Association for Research in Science Teaching



(USA) and the British conference on International Comparisons of Pupil Performance: Issues and Policy, held at the University of Leeds. Despite the prestige of the overseas occasions, the conferences with the most potential for IEA studies to influence Australian education are the domestic ones, including those run by the subject associations, the Australian College of Educators and university faculties, together with invited talks to curriculum authorities and education departments, at which many ACER researchers have presented papers.

Further dissemination strategies designed to reach as far as possible into school systems and head offices are the articles on IEA studies that appear in ACER's newsletter, *Research Developments*, and articles in *Education Review* and the *Professional Educator*, both being produced by the Australian College of Educators. TIMSS results are also regularly mentioned in the reports of the Federal Productivity Commission.

### **The trend studies in the fields of mathematics and the sciences**

The Third International Mathematics and Science study (TIMSS) was the largest and most ambitious comparative study ever undertaken, with more than half a million student participants from three stages of their schooling from 45 countries around the world. The study followed the First and Second International Mathematics Studies (1963-1967, 1976-1987) and both IEA Studies in Science (1970-1971, 1983-1984), and brought together the mathematics and science assessments under one umbrella. Australia, through the Australian Council for Educational Research, participated in the first of the international mathematics studies and a repeat study in 1978, and both of the science studies. Australia's formal involvement in TIMSS began in late 1991, when a proposal was presented to, and supported by, the Australian Education Council.

TIMSS broke new ground in Australia in terms of the scope of its operation and its test materials, curricular analysis and the contextual information collected. At the end of 1994, ACER tested students at

three different population levels: middle primary, junior secondary and upper secondary school, and from 1996 to 1999 produced a report in three volumes examining performance in mathematics and science in Australian schools (Lokan, Ford & Greenwood, 1996, 1997, 1999). In 1998, Australian secondary schools participated in a partial replication of TIMSS at Year 8 level, and in 2002, with widespread participation and funding secure for the study, TIMSS became the Trends in International Mathematics and Science Study, with an added emphasis on monitoring trends, and focused on two year levels only – Year 4 and Year 8. Subsequent to this, Australia has participated in TIMSS 2007 and is currently involved in TIMSS 2011.

Australia, through ACER in each case, has continued its strong involvement not only with the implementation and reporting of TIMSS in Australia but has also been closely involved in terms of item development and in the development and review of the contextual questionnaires. Through the National Advisory Committee, feedback on data requirements have been readily available from governments, practitioners and others with a vested interest in education, and these have helped form the reporting strategy of the National Centre at ACER. These reports focus on the achievement, attitudes and aspirations of Australian students in comparison to those internationally. They also provide a detailed examination of the results by State and for the socio-demographic groups that are currently of particular government concern: gender, Indigenous students, geographic location, language background and parental education (as a proxy for socioeconomic background).

There is always great anticipation for the TIMSS results, largely with respect to where Australia is ranked internationally, whether Australia has done better or worse than in previous studies, and whether other countries have improved their scores. The media are always interested in contextual items such as the reported level of bullying in Australian schools since Australia and New Zealand have lower than average rates for students feeling safe at school. There is an initial flur-

ry of enquiries from and presentations to Commonwealth, State and Territory Government Departments followed by often more detailed queries. The proportion of students achieving at or above the intermediate benchmarks at both year levels, in mathematics and science, is published by the Productivity Commission.

As well as the national reports published at the end of each cycle (Thomson & Fleming, 2004a, 2004b; Thomson, Wernert, Underwood & Nicholas, 2008), the research team at ACER use the national data in a number of different ways in order to inform a wide audience. For example, to facilitate government discussion on Indigenous Education, an in-depth analysis was conducted of the TIMSS 2003 results for Indigenous students (Thomson, McKelvie & Murnane, 2006). This level of reporting was possible because ACER deliberately over-sampled Indigenous students in all international sample studies, providing sufficient numbers for more detailed analysis than would be possible if the only Indigenous students were those who were sampled as part of the random sample.

After the TIMSS 2006 cycle, two particular reports were published that examined the curriculum aspect of the TIMSS project. One focusing on mathematics and one on science, the reports examined Australian students' achievement on the released TIMSS items – some on which most Australian students did well and some not so well. All released items were included with the reports, along with marking guides, on CD.

After TIMSS 2006, ACER was invited to work with staff from the Victorian Department of Education and Early Childhood Development (DEECD) to examine the TIMSS results from Victoria and New South Wales in order to attempt to identify reasons why New South Wales' performance on TIMSS was better than that of Victoria. This unpublished work formed part of a review into mathematics and science education in Victoria.

The methods used in TIMSS have influenced the ways in which sample studies in the National Assessment Program in Australia have

been conducted. In particular the rotated block designs used in TIMSS in 1994-95 to enable breadth of curriculum coverage have been followed, as have the two-stage cluster sample designs, applied at specified grades, developed through TIMSS.

### *Probing Mathematics and Science Teaching Practices More Deeply*

#### **The pilot TIMSS video study**

IEA, through the so-called 'video studies' undertaken as part of TIMSS and the repeat of TIMSS carried out four years later (TIMSS-R), introduced videotaping of national samples of mathematics and science lessons in an attempt to obtain better information about teaching practices. The feasibility of capturing data in this way on this scale, and analysing the data to produce meaningful results, was assessed during TIMSS when a pilot study of eighth grade (Population 2) mathematics lessons was undertaken in Germany, Japan and the United States (Stigler & Hiebert, 1999). The United States had performed poorly in SIMS, leading to considerable interest in going beyond the achievement data to understand as much as possible about the reasons why. Japan was chosen for comparative purposes because its students had always been among the highest achievers in international studies; and Japan and Germany, as well as being significant economic competitors of the United States, were both considered to have research teams that would be able to implement the study.

It has long been assumed that what teachers did during class lessons was related, probably causally, to what students achieved. IEA studies had typically included teacher questionnaires, but were limited by what could be asked in a questionnaire requiring no more than an hour to answer, since any longer would jeopardise response rates. There was also the concern that respondents might well interpret questions differently: for example, what one teacher might regard as 'problem-solving', another might regard as 'routine exercises'. From

questionnaire responses, there was no objective way of knowing what had actually occurred in class lessons.

Videotaped class lessons have been used for many years as an instructional aid in teacher education and for small-scale qualitative studies. However, their use on a wide scale for cross-cultural research in TIMSS in the mid-1990s was highly innovative. So that comparisons could legitimately be made, many methodological hurdles had to be overcome to ensure that the filming was done in a standard way in all classrooms. Thus, samples had to be selected to be representative, and very detailed coding schemes had to be developed that could be used reliably by teams of coders. In developing these schemes, the researchers had to keep in mind that the codes should be able to be used, singly or in combination, to describe teaching in useful ways. All of the processes in the study were labour-intensive, requiring a much higher level of funding than for the paper-and-pencil assessments typical of cross-national studies. Support for the pilot video study was provided by the National Center for Education Statistics (NCES) of the Department of Education in the United States.

The pilot video study involved 231 classrooms across the three countries, 100 in Germany, 81 in the United States and 50 in Japan, selected randomly from within the larger sample of classes that took part in the TIMSS written assessment. Budget constraints meant that each class could be filmed only once. Filming took place fairly uniformly across the school year in Germany and the United States, but this was not achieved in Japan. This was a problem for the study in that their curriculum was followed in a uniform way across the country as the year progressed. Thus, most classes filmed in Japan involved geometry topics. Despite these shortcomings, enough intelligible differences between teaching methods were found among the three countries to encourage including video studies in TIMSS-R that involved more countries and examined science classrooms as well as mathematics classrooms, again at the eighth grade.

### **The more extensive TIMSS-R video studies**

Seven countries – Australia, the Czech Republic, Hong Kong Special Administrative Region (hereafter referred to as Hong Kong), Japan, The Netherlands, Switzerland and the United States – took part in the video study of mathematics classrooms as part of TIMSS-R. Students in all of these countries had scored significantly above their counterparts in the United States in TIMSS and TIMSS-R. Hong Kong and Switzerland chose not to be involved in the Science study, leaving five countries as the participants in that component. Again, the study was financed substantially from the United States, as a main focus for them remained to identify teaching practices characteristic of instruction in higher achieving countries that were not typical of instruction in the United States. Within Australia, overseas funding was supplemented by the Commonwealth and State Governments and the ACER, which managed the study locally, so that an ACER staff member could be based at LessonLab in Los Angeles which managed the study overall. Hilary Hollingsworth was thus able to contribute to the code development and data analysis phases, so that full Australian reports of both the mathematics and science components could be prepared (Hollingsworth, Lokan & McCrae, 2003; Lokan, Hollingsworth & Hackling, 2006).

### **Sampling**

National representative random samples of lessons were selected for filming in each country, usually in the same schools that had carried out the written assessment in 1999, but in Australia a new school sample was needed because the written assessment took place in 1998. The new sample was selected in the usual manner, stratified by State and sector, and with potential replacement schools selected at the same time in case of refusals. A methodological feature to be dealt with was that replacement lessons were not permitted, so that if a teacher refused to take part, another school had to be selected. In Australia this was managed by first having the principals of the selected schools dis-

cuss the project with their teachers, advising them that the video researchers would appear to film a class with very little prior notice of no more than a day or two. Some of the schools selected originally did not wish to participate, but with the use of replacement schools the weighted Australian response rate was 87 percent, or 87 from a designed sample of 100 classes, one from each of 100 schools. The mathematics lessons were a truly random sample, but, for practical reasons because of distance, the science lessons selected for filming were chosen randomly from a list of those occurring on the same day as, or on one day either side of, the selected mathematics lessons. Altogether 638 eighth grade mathematics lessons from seven countries and 439 science lessons from five countries were filmed for the study, spread as evenly as possible throughout the school year.

### **Other methodological aspects**

The IEA video studies undoubtedly contributed a great deal to the development of methods employed for this form of research. Two high quality cameras, provided by LessonLab, were used in each classroom in the larger study, one at the side or back of the room and focused on the whole room, the other with a detailed protocol for focusing on specified aspects of the lesson, depending on how the lesson unfolded. A great deal of work went into development and refinement of the specifications for the videotaping; for example, when to follow the teacher, when to look at what individual students or groups of students were doing, and when to show a close-up of a piece of apparatus or a worksheet. Videographers then had to be found, trained and supervised while taping some trial lessons that were not part of the study, to ensure that they were correctly interpreting the instructions and procedures. For some countries the training took place at LessonLab, but for others, including Australia, it was carried out by travelling LessonLab staff.

Some prior work from the pilot study was available on coding schemes for quantifying the mathematics data. Further work for the

1999 study was carried out over about two years by a team of mathematics educators and psychologists in order to extend and then refine the codes. Trials had shown which codes were not reliable enough for larger groups of coders to apply and which codes were likely to be the most useful as descriptors of the data for analysis purposes. A further two years were then spent on producing codes for the science data. All coding of data was undertaken by country representatives and project staff at LessonLab, with regular monitoring to ensure that the codes were applied as reliably as possible across countries. All of these processes, including the sampling and weighting procedures, were described in great detail in the international technical reports from the studies (Jacobs, et al., 2003; Lemmens, Druker, Garnier, Chen & Roth, 2006). There is also a Technical Appendix in each of the two Australian reports (Hollingsworth, Lokan & McCrae, 2003; Lokan, Hollingsworth & Hackling, 2006).

### **Implications of the findings for Australian teaching and curricula**

Findings from the video studies were used to ascertain the extent to which Australian teaching reflected curriculum and other documents developed during the 1990s. The studies sought to view Australian teaching practices in comparison with those in some of the countries with the highest levels of student achievement, particularly Japan, and to obtain exemplars of teaching in mathematics and science that could be used for subsequent research and professional development purposes. The science results, which were more substantially positive and therefore led to less concern among Australian educators, are discussed first.

#### *Science*

The report of the Science Study (Lokan, Hollingsworth & Hackling, 2006) yielded mostly very positive results for Year 8 science teaching in Australia. Most of the lessons took place in laboratories, which were well equipped. The only resources teachers said they were short of



were computers and internet access. This was a concern in 1999 but should not be a concern a decade later. On average, less than two percent of the lesson time was spent on matters unrelated to science instruction. Nevertheless, almost half the lessons experienced an 'outside interruption', for example a broadcast message, compared with only seven percent of lessons in the Czech Republic and none in Japan. New content was introduced and discussed in all but three lessons, occupying 85 percent of the time on average. The level of challenge of the scientific content was similar in all countries except the Czech Republic, judged to be 'basic' in about half the lessons. The lessons were assessed as 'basic' in 57 percent of the Australian lessons but in only 18 percent of the Czech lessons; 'challenging' in about 10 percent of the lessons except for 19 percent in the United States and 25 percent in the Czech Republic. Real-life issues and first-hand data were often used in the well-structured lessons to support the development of ideas, with a variety of hands-on activities likely to engage students' interest. Australian teaching was found to be remarkably similar to Japanese teaching in many respects, particularly in the extent of practical or seatwork activities undertaken by students independently of the teacher and in the way that conceptual links were used in the material presented during content-focused lessons.

In keeping with one of Australia's goals for taking part in the TIMSS-R video studies, the Australian report of the science component contained an analysis of the goals of school science and characteristics of effective science teaching gleaned from recent research literature. Based on these, and the rationale that "...developing scientific literacy should be the focus of science education in the compulsory years of schooling" (Hackling, Goodrum & Rennie, 2001, p.6-7), an ideal picture of science education was constructed. Measured against this ideal scenario, as presented in the report, the Australian picture of typical science teaching at Grade 8 revealed by the results from the science component of the video study provided strong endorsement for the quality of that teaching overall. Some aspects requiring improvement

were identified, namely that students had little opportunity to: (a) formulate their own research questions, (b) design their own investigations, or (c) make predictions, thereby limiting their opportunities to develop higher order skills.

### *Mathematics*

In contrast to science, there were few very positive outcomes of the mathematics video study for Australia, although Australia was not alone in this regard. Some positive aspects were that over 70 percent of the Australian teachers said that they believed their teaching reflected recent developments in mathematics education in their country, exceeded only by the United States teachers (86 percent). More than a quarter of the Australian teachers used real-life contexts to introduce material, and Australia, as well as Japan and Switzerland, were the only countries in which real-world objects were used in more than a fifth of the lessons. Students used calculators in over half the lessons in Australia and Switzerland, exceeded only by lessons in The Netherlands (91 percent). Significantly more lesson time, on average, was devoted to introducing and practising new content in Australia than to reviewing previous content, as it was in all countries except the Czech Republic and the United States.

Closer examination of classroom events revealed shortcomings for most countries in one or more aspects of their mathematics teaching, and many were found for Australia. The Australian lessons were characterised by low-level mathematics content. Problems typically involved procedures rather than reasoning. The procedural level of the problems was generally low, as it was in most countries, and the problems were typically of short duration and presented in repetitive sets. Over half (55 percent) of the Australian lessons, from a sub-sample randomly selected for the purpose, were judged to include content at not even a moderate level of complexity, compared with 5 percent in Hong Kong, 15 percent in the Czech Republic, 30 percent in The Netherlands and Switzerland and 45 percent in the United States. The incidence of

problems worked on for more than 45 seconds per problem was higher in Hong Kong, Japan, The Netherlands and Switzerland than it was in Australia, the Czech Republic and the United States.

A further opportunity to probe mathematics teaching arose in Australia through comparisons of video study results from higher- and lower-achieving classes. A subset of the TIMSS-R mathematics test was given to the video study participants so that these comparisons could be made. Most of the findings from these comparisons yielded expected results, for example: (a) more time was spent in the higher-achieving than the lower-achieving classes on new content, (b) students spent more time working on longer problems and working independently of the teacher, (c) more time was spent in the lower-achieving than the higher-achieving classes working on problems in groups, with more teacher input, and (d) more time was spent on review. An unexpected, and disturbing, result was revealed in that there was little difference between the contrasted groups of classes regarding the level of complexity of the problems the students worked on, regardless of whether the students were working concurrently on the same problems or whether students were working independently at their own pace.

Some of Australia's best-known mathematics educators provided comments in the Australian report, which have strong implications for mathematics curricula and teaching. The writers identified a syndrome of shallow teaching, relying heavily on procedures without reasons, excessive repetition of many short and similar exercises, too much use of problems of low complexity, only average use of real-world contexts and little evidence of motivating activities. One of the mathematics educators wrote of the video studies (Hollingsworth, Lokan & McCrae, 2003, p. 106),

If these videos and data represent fairly normal current practice in these countries (and the teachers involved say they do), then there are a lot of pretty boring, artificial, low-level, irrelevant, mentally stifling lessons being delivered [...] in the name of Year 8 mathematics, and it is not surprising that so many adults don't want to

know anything more about mathematics after they leave school.

The implications for Australia are that our students need to be challenged more in mathematics, especially our more able students, giving them more opportunities for working on longer, more complex problems and for developing higher-order skills such as mathematical reasoning, deduction and making connections.

### **Comment**

The artificiality of having cameras present during classroom lessons, especially as one of the videographers moved around the room and was therefore more noticeable, was often mentioned as a threat to the validity of the observations derived from filming the lessons. Data from the teacher questionnaires showed that the threat was not a major factor in this study for Australia. This was also found in most countries, except for the Czech and Japanese mathematics teachers, 35 percent of whom said their teaching was worse than usual. How much this might have affected the results, of course, could not be determined in these two countries.

In both mathematics and science, at least three-quarters of the Australian teachers (80 percent in science) said that their videotaped lesson was typical or very typical of their teaching and over 95 percent said their students' behaviour was about the same as or better than usual (22 percent in mathematics and 27 percent in science answered 'better'). Three-quarters of the science teachers and 80 percent of the mathematics teachers (where the videographers did not need to move around the room as much) said that the presence of the cameras did not affect the nature or quality of their teaching. Of the remaining teachers, twice as many, in both mathematics and science, said that their teaching was worse than usual than said it was better than usual. A few teachers seemed nervous as their lesson began, but the tapes showed that this quickly decreased as the teachers and students settled into their routines.

### *Information and Communication Technology in Australian Schools*

In Australia, the incorporation of Information and Communication Technology (ICT) into school programs has been steadily progressing since the late 1980s. Australia is now characterised as a country in which students and teachers make relatively extensive use of ICT (Anderson & Ainley, 2010). The Australian National Goals for Schooling formulated by the Ministers for Education in 1999 asserted that when students left school they should be: “confident, creative and productive users of new technologies, particularly information and communication technologies, and understand the impact of those technologies on society” (Ministerial Council on Education, Employment, Training and Youth Affairs (MCEETYA, 1999, “National Goals” para. 1). Nine years later the importance of ICT in education was reiterated in the Melbourne Declaration on Educational Goals for Young Australians (MCEECDYA, 2008, p. 04-05) which stated that “in this digital age young people need to be highly skilled in the use of ICT”. More recently times have seen the implementation of the ‘digital education revolution’ as a feature of education policy. This national initiative involves significant support for improving ICT provision in schools, expanding the use of ICT in teaching and learning and developing the ICT proficiency of young Australians.

Two of the three modules of the IEA Second International Technology in Education Study (SITES) were conducted in Australia. Both were concerned with how ICT was used in classrooms to support innovative approaches to teaching and learning. SITES Module 2 was a study of innovative pedagogical practices using technology involving 174 case studies in 28 countries (Kozma, 2003). In each country there was a systematic process for selecting the cases and across all the cases there was common protocol and a common set of instruments used in the research. Australia contributed five cases to the international study and an additional four cases for its own national analyses (Law, Pelgrum & Plomp, 2008). The five Australian cases contributed to the

international study included: (a) a whole school approach to integrating ICT across the curriculum (Queensland), (b) using multi-media development tools to foster learning (South Australia), (c) an online secondary school orientation program (Tasmania), (d) embedding ICT into an integrated English and history project (Victoria), and (e) using electronic Distance Education to extend opportunities for secondary students (Queensland). One of the additional cases involved ICT used in a predominantly Indigenous school.

One of the challenges of a study such as SITES Module 2 was handling the qualitative data from 174 cases. The approach was to blend narrative accounts of cases with a cluster analysis process that identified common patterns in the coded data from the cases. SITES Module 2 identified seven patterns of innovation using ICT (Kozma, 2003), (a) communication and productivity tool use, (b) collaborative student research involving groups collecting and analysing data, (c) information management focusing on searching for, organising, managing and using information, (d) teacher collaboration on design of instructional materials and activities, (e) outside communication that involves students working with others outside the classroom (using email, the internet, conferencing software and listservers), (f) product creation that focuses on the design of digital products, and (g) tutorial projects in which software is used to provide opportunities to practise and refine skills.

The third of the SITES modules was designated as SITES 2006 and was a survey of teachers of junior secondary school mathematics and science, along with their schools. The major source of data was SITES 2006 (Law, Pelgrum & Plomp, 2008) which included, in total across 23 countries, a little fewer than 31,000 teachers from almost 8,300 schools. Australia participated in SITES 2006 as a benchmarking country conducting its national survey in 2007. In Australia, the achieved sample was just under 1100 teachers from 301 schools.

Analyses of SITES data indicated that Australian science and mathematics teachers were relatively high users of ICT. Use of ICT was

greater when teachers had a higher level of confidence (or self-reported competence) in ICT, when there were fewer contextual obstacles (infrastructure, digital learning resources, ICT access) and when there was a relatively higher level of computer provision in schools (Ainley, Eveleigh, Freeman & O'Malley, 2011; Ainley, 2010; Ainley, Banks & Fleming, 2002). In Australia, the teaching of Grade 8 Science was more likely to make use of ICT than the teaching of Grade 8 Mathematics--a difference that was also evident in many other education systems. The samples of mathematics teachers and science teachers were taken from the same schools and so the differences could have reflected curriculum differences at system level in these areas or differences in the nature of the disciplines. It may be that some subjects lend themselves more readily to the pedagogical use of ICT or that there are stronger traditions of innovation in some subjects than others.

Comparative international studies such as SITES provided a context for national perspectives on educational issues such as the use of ICT in teaching. When data from SITES in Australia were compared with data from other countries they suggested that ICT had been relatively widely adopted, that there was a relatively strong provision of computers in schools, and teachers were more confident in their ICT capability than their peers in other countries. However, the data also suggested that the implementation of ICT in teaching could be enhanced by building the capacities of teachers as well as improving the resources available to students and teachers.

The next phase of IEA studies targeting ICT in education is planned to have a focus on students and computer-based assessments of ICT proficiency among students as well as gathering information from teachers and schools about how ICT is used. The study will be known as the International Computer and Information Literacy Study (ICILS) and is scheduled to collect data in 2013. The Australian Council for Educational Research will be the International Study Centre for ICILS and will use its experience of the national assessment of ICT Literacy to inform the instruments to be employed (MCEECDYA, 2010).

### *Development of Measurement in Educational Research: Rasch Scaling*

Perhaps the most successful development arising from the IEA program of research in Australian education is that associated with the evaluation of educational achievement and assessment of scholastic aptitude. The IEA centre located in the Department of Education at the University of Chicago under the guidance of Ben Bloom and Ben Wright during the 1960s and 1970s advanced the expansion and redevelopment of programs of educational evaluation at the Australian Council for Educational Research, and across Australia. Subsequently, the work undertaken in Australia has been fed back overseas by Barry McGaw, David Andrich, Ray Adams, Mark Wilson and Geoff Masters to influence the fields of educational assessment and evaluation worldwide.

The visits of Ben Bloom and Ben Wright to the ACER in the 1970s and the Conference held in 1980 to mark the 50 years from the foundation of the ACER in the 1930s, and the papers (Spearritt, 1982) given by Bob Thorndike and Bruce Choppin served to inspire and augment the work into Rasch scaling that had already been initiated. The flow of young scholars from Australia to Chicago over the years, who included Ray Adams, David Andrich, Michael David, Graham Douglas, Geoff Masters, George Morgan and Mark Wilson, built teams of research workers in the field of Rasch measurement who have subsequently undertaken further advances over the past three decades. Strengths have been developed in four distinct fields of inquiry and endeavour, namely: (a) monitoring of educational achievement, (b) test equating, (c) the development of measurement models, and (d) the reporting of educational performance.

#### **Monitoring of educational achievement**

The first important advance in an educational context during this period of a third of a century was an increasing concern for educational accountability and a growing interest in monitoring achievement and the reporting of student performance to the students, parents, teachers



and the wider public. The first major national survey of literacy and numeracy performance that occurred in 1975 was the Australian Studies of School and Student Performance (ASSP) (Bourke & Keeves, 1976; Keeves & Bourke, 1977). This national survey was repeated in 1980 with assessment at the 10-year-old and 14-year-old age levels (Bourke, Mills, Stanyon & Holzer, 1981). These surveys provided baseline data in literacy and numeracy. Subsequently, one by one, each of the States introduced educational assessment programs to test and report on the achievements of entire cohorts of students. By the turn of the century, in the year 2000, all States and Territories had introduced tests for all students at the levels of Year 3 and Year 5, usually in both literacy and numeracy.

In parallel with these State and Territory assessment programs, Australia undertook a National School English Literacy Survey in 1996 to examine the reading, viewing, speaking, listening, writing, and spelling skills of Year 3 and Year 5 students (Masters & Forster, 1997a, 1997b). Australia also participated at all three levels (primary, junior secondary and upper secondary) in the Third International Mathematics and Science Study (TIMSS) in 1994 (Lokan, Ford & Greenwood, 1996, 1997, 1999) and took part in the TIMSS-Repeat study and the IEA Civics Study in 1998. During the first decade of the twentieth century these programs were extended to cover complete cohorts of students at Years 3, 5, 7 and 9, in literacy and numeracy on an annual basis, together with intermittent testing programs in specific curricular fields. Gradually Rasch measurement procedures were introduced into all testing programs, leading to greater rigour and improved standards of reporting.

### **Test equating**

The First and Second IEA Science Survey testing programs contained sufficient common items for the effective calibration of the tests using Rasch scaling (Keeves, 1992a). This led to the formation of a common scale and absolute changes over time could be measured. Thus trends

in student achievement could be assessed within each of the countries involved. Under the supervision of Ray Adams at the ACER, Australian researchers calibrated the mathematics and science items in the Third International Mathematics and Science Study, effectively equating the different forms of these tests across time and grade levels and across the participating countries. This was the largest international survey undertaken and possibly the most complex set of objective measurement analyses that had ever been carried out (Lokan, Ford & Greenwood, 1996, 1997, 1999).

This work not only set the standards for all testing programs that were conducted within Australia, but also set the standards for the testing programs conducted by the Programme for International Student Assessment (PISA) within the Organisation for Economic Cooperation and Development (OECD). The contract for the management of the PISA study in the year 2000 was awarded to an international consortium led by Australian measurement researchers at the Australian Council for Educational Research. Other similar contracts for the scaling of IEA data followed. Subsequently, Australian testing programs have maintained the use of Rasch scaling procedures. Moreover, Andrich from the University of Western Australia has supervised the work led by former ACER staff members Ken Ross and Hungi Njora at the International Institute for Educational Planning in Paris for the Southern and Eastern African Consortium for Monitoring Educational Quality (SACMEQ) in Southern Africa.

### **The development of measurement models**

Measurement in education is underpinned by a family of statistical models, known as 'conjoint measurement models'. The first of these models was advanced in Edinburgh in the early 1940s, and subsequently developed by George Rasch using the logarithmic function as an alternative to the normal distribution. During the 1980s and 1990s Australian educational researchers were among the most active in constructing, describing and scaling variables for use in educational set-

tings to advance measurement in the different fields of education and educational research. Pioneering work was undertaken by Andrich (1978a; b; 1997) for the measurement and analysis of ratings; Geoff Masters (1997) for the assignment of partial credit; Mark Wilson in the development of the ordered partition model (1992) and the Saltus model (Wilson, 1989), and by Ray Adams and Mark Wilson for the multidimensional item response models (Adams, Wilson & Wang, 1997a) as well as the multilevel item response models (Adams, Wilson & Wang, 1997b). These Australian measurement research workers made important theoretical advances and provided new measurement procedures and supporting software that included the programs known as MLTBIN, RATE, DISLOC, RUMM, RUMMFOLD, CREDIT, QUEST and CONQUEST.

### **Reporting education performance**

The use of measurement theory and Rasch scaled scores in large-scale testing programs led to the introduction of more meaningful methods of reporting student, school and system level outcomes to the many and various stakeholders in the provision of educational services. Pictorial representations of change in performance over time, together with indicators of the range of error involved in measurement, replaced raw scores or percentage scores. This was possible because measures of performance were recorded on a scale of measurement that remained unchanged over time and that was not truncated at the upper and lower levels, thus allowing for the assessment of growth and development. An early example of the reporting of state-wide test results recorded on calibrated measurement scales was ACER's work on the 1989 New South Wales Basic Skills Tests (Masters et al., 1990). In 1992, this work, supervised by Jan Lokan, was awarded the triennial United States National Council of Measurement in Education Award for the dissemination of educational measurement concepts. Similar forms of reporting were adopted across Australia, and by 1997 government educational authorities required the use of objective

measurement theory as the basis for the analysis and reporting of system-wide test results.

These developments could not have taken place across Australia to transform the programs of assessment and evaluation of student performance at school on scales of measurement in an objective way that indicated change and growth over time, without the support of at least one university in each State conducting a course in the theory and practice of measurement of educational outcomes. Furthermore, the use of scholastic aptitude tests, using Rasch measurement scales, for university entry and selection, has spread across Australia and subsequently overseas to Great Britain, the Middle East and to Southeast Asia under the guidance and supervision of ACER.

### **Conclusion**

The Australian Council for Educational Research was established in 1930 to provide "stimulation from within" to the State educational systems in Australia (Connell, 1980, p. i). This was to be carried out through the promotion, "as far as possible in cooperation with existing institutions, the cause of research and investigation in Australia" (Connell, 1980, p. ii). The support for the first 15 years of the life of the Council came from the Carnegie Corporation of New York that had been established in 1911 to "promote the advancement and diffusion of knowledge" (Connell, 1980, p. iii) and a proportion of its resources were set aside as a special fund for use in Canada and the British colonies. Thus, the involvement of the ACER in the research programs of the IEA was fully consistent with the purposes for which the Council was formed, since in 1934 the ACER had been made a central bureau for Australian education by the League of Nations Committee on Intellectual Cooperation.

No formal examination of the impact of the IEA program of research on the ACER and through the ACER on education in the Australian

States and Territories has been undertaken. However, Connell (1980, p. iv), in writing a review of the history of the ACER over the first 50 years of its operation, made the following comments that covered the period from 1963 through to 1979 on the influences of the IEA on the Council and on Australian education:

The IEA program received a considerable amount of attention at the ACER, from the time when the first tentative connection was made in 1963, and the staff was not always sure that it was a worth-while expenditure of time [...] Australian educators, however, had something to learn from the expert curriculum analysis on which the testing program was based, and research workers, from the sophisticated technique of sampling and analysis that characterized the project. But some of the staff, seeing the program as essentially one of monitoring student performance, wished for some more creative research that would investigate processes and not merely results. If the IEA did not do that, it did at least make a solid contribution to an understanding of the conditions of learning. It took up the task of explaining the influences which had important effects on a student's score and, in particular, sought evidence, through its questionnaires on students and teachers, on the currently vexed problem of the extent to which school circumstances affected performance. Much of what produced differences in performance could not be explained, but it could be shown that the cultural and economic circumstances of the home background and the condition and organization of the school were two important influences. Despite recent views that school circumstances were not of great significance, studies of the IEA material in general, and special studies of the material on the Australian tests in science, were able to show that they were of considerable importance and to indicate the particular aspects of schooling that needed attention if students' performances were to be improved.

Thirty years have passed since the substantial statement recorded above was written. During that time ACER has become more heavily

involved in IEA activities that have broadened in scope, and together with educational research, in general, have increased in strength. At the same time research activity, like higher education, has become increasingly commercialized, and, like the Australian universities and higher educational institutions worldwide, ACER has entered the marketplace (Bok, 2003). As a consequence the ACER has become free to expand its program of research and development. This has provided opportunities to open up new fields of inquiry, particularly in educational measurement and the conduct of survey research studies. ACER has, as a result, won substantial contracts that were based on and directly derived from its affiliations and experience obtained through its involvement in IEA. It is perhaps premature to make an assessment and evaluation of this expansion in activity, since the Australian State and Territory systems of education are engaged in a so-called 'revolutionary' phase that is led by the Australian Government. The material presented in this chapter records in its later sections the advances made in ACER's involvement in the IEA research program and has sought to show the ways in which these activities are influencing Australian education. The danger that both the ACER and the IEA face is that they have entered a phase in which they must both respond to an agenda set by others from an overseas base and perspective, rather than do what both were able to do in their initial years of operation, namely to lead and stimulate change for the betterment of educative processes worldwide.

Nevertheless, during recent years, the ACER has become a large and highly respected educational research organization not only within Australia but also in a worldwide setting. These developments have taken place largely as a direct consequence of the involvement of the ACER in the wide range of IEA's studies and activities in educational research and the evaluation of educational achievement across a growing number of countries of the world.

## References

- Adams, R. J., Wilson, M. R., & Wang, W. C. (1997a). *The Multidimensional Random Coefficients Multinomial Logit*. Melbourne, VIC: ACER.
- Adams, R. J., Wilson, M. R., & Wang, W. C. (1997b). Multilevel item response models: An approach to errors in variables regression. *Journal of Educational and Behavioural Statistics*, 22 (1), 47-76.
- Ainley, J. (2010). Influences on the implementation of ICT in teaching in Australia. Paper Presented at the IEA International Research Conference, Gothenburg, Sweden.
- Ainley, J., Banks, D., & Fleming, M. (2002). The influence of IT: Perspectives from five Australian schools. *Journal of Computer Assisted Learning*, 18 (4), 395 -405.
- Ainley, J., Eveleigh, F., Freeman, K., & O'Malley, K. (2011). *ICT in the teaching of Science and Mathematics in Year 8 in Australia: A report from the SITES survey*. Melbourne, VIC: ACER.
- Anderson, R. & Ainley, J. (2010). Technology and learning: Access in schools around the world. In B. McGaw, E. Baker & P. Peterson (Eds.), *International Encyclopaedia of Education* (3rd edition). Amsterdam: Elsevier.
- Andrich, D. (1978a). A Binomial Latent Trait Model for the Study of Likert-style Attitude Questionnaires. *British Journal of Mathematical and Statistical Psychology*, 31 (1), 84-98.
- Andrich, D. (1978b). Scaling Attitude Items Constructed and Scored in the Likert tradition. *Educational and Psychological Measurement* 38(3), 665-680.
- Andrich, D. (1997). Rating Scale Analysis. In J. P. Keeves (Ed.), *Educational Research, Methodology and Measurement: An International Handbook* (2nd edition), (pp. 874-880). Oxford: Pergamon.
- Australian Council for Educational Research (ACER) (1997). *Raising Australian Standards in Mathematics and Science: Insights from TIMSS*. Conference Proceedings. Melbourne, VIC: ACER.
- Baker, D. P., & Jones, D. P. (1993). *Creating Gender Equality, Cross-*

- national Gender Stratification and Mathematical Performance. *Sociology of Education*, 66 (2) 91-103.
- Bloom, B. S. (Ed.). (1956). *Taxonomy of Educational Objectives: The Classification of Educational Goals. Handbook 1. Cognitive Domain*. New York: McKay.
- Bloom, B. S., Hastings, J. T., & Madaus, G. F. (1971). *Handbook of Formative and Summative Evaluation of Student Learning*. New York: McGraw-Hill.
- Bok, D. C. (2003). *Universities in the Marketplace: The Commercialization of Higher Education*. Princeton, NJ: Princeton University Press.
- Bourke, S. F. (1984). *The Teaching and Learning of Mathematics*. Hawthorn, VIC: ACER.
- Bourke, S. F., & Keeves, J. P. (1976). *Australian Studies in School Performance Volume I: Literacy and Numeracy in Australian Schools, a First Report*. Canberra: Australian Government Publishing Service.
- Bourke, S. F., Mills, J. M., Stanyon, J., & Holzer, F. (1981). *Performance in Literacy and Numeracy 1980*. Canberra: Australian Government Publishing Service.
- Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical Linear Models: Applications and Data Analysis Methods*. Newbury Park, CA: Sage.
- Carroll, J. B. (1963). A model of school learning. *Teachers College Record*, 64(8), 723-733.
- Comber, L. C., & Keeves, J. P. (1973). *Science Education in Nineteen Countries*. Stockholm: Almqvist and Wiksell.
- Connell, W. F. (1980). *The Australian Council for Educational Research 1930-80*. Hawthorn, VIC: ACER.
- Darmawan, I. G. N., & Keeves, J. P. (2006). Suppressor Variables and Multilevel Mixture Modelling. *International Education Journal* 7(2), 160-173.
- Doran, R. L., & Tamir, P. (Eds.). (1972). Practical Skills Testing. *Studies in Educational Evaluation*, 18 (3), 263-406.
- Fordham, A. M. (1983). *The Context of Teaching and Learning*. Hawthorn, VIC: ACER.



- Fraser, B. J. (1997). Classroom Environments (pp. 896-900); Descriptive Scales (pp. 900-908). In J. P. Keeves (Ed.), *Educational Research, Methodology, and Measurement. An International Handbook*. (2nd ed). Oxford: Pergamon.
- Hackling, M. W., Goodrum, D. & Rennie, L. (2001). The State of Science in *Australian Secondary Schools. Australian Science Teachers Journal*, 47(4), 6-17.
- Hollingsworth, H., Lokan, J., & McCrae, B. (2003). *Teaching Mathematics in Australia: Results from the TIMSS 1999 Video Study* (TIMSS Australia Monograph no. 5), p.106. Melbourne, VIC: ACER.
- Husén, T. (Ed.). (1967). *International Study of Achievement in Mathematics* (2 vols). Stockholm: Almqvist and Wiksell.
- IEA (1988). *Science Achievement in Seventeen Countries. A Preliminary Report*. Oxford: Pergamon.
- Jacobs, J., Garnier, H., Gallimore, R., Hollingsworth, H., Givvin, K. B., Rust, K., Kawanaka, T., Smith, M., Wearne, D., Manaster, A., Etterbeek, W., Hiebert, J. & Stigler, J. W. (2003). *TIMSS 1999 Video Study Technical Report, Volume 1: Mathematics Study* (NCES 2003-012), US Department of Education. Washington, DC: National Center for Education Statistics.
- Jones, W. B. (1988). *Secondary School Mathematics and Technological Careers. Hawthorn, VIC: ACER*.
- Keeves, J. P. & Radford, W. C. (1969). *Some Aspects of Performance in Mathematics in Australian Schools*. Hawthorn, VIC: ACER.
- Keeves, J. P. & Read, A. D. (1974). *Sex Differences in Preparing for Scientific Occupations* (IEA Australia Report 1974:6). Hawthorn, VIC: ACER.
- Keeves, J. P. (1968). *Variation in Mathematics Education in Australia*. Hawthorn, VIC: ACER.
- Keeves, J. P. (1972). *Educational Environment and Student Achievement*. Stockholm: Almqvist and Wiksell.
- Keeves, J. P. (1973). Differences between the Sexes in Mathematics and Science Courses. *International Review of Education*, 19(1), 47-62.
- Keeves, J. P. (1974). The IEA Science Project. In K. Frey (Ed.), *Implementa-*

- tion of Curricula in Science Education*, pp. 158-176. Köln, West Germany: Deutsche UNESCO-Kommission.
- Keeves, J. P. (1995). *The World of School Learning*. The Hague: IEA.
- Keeves, J. P. (1992b). *Learning Science in a Changing World*. The Hague: IEA.
- Keeves, J. P. (Ed.). (1992a). *The IEA Study of Science III: Changes in Science Education and Achievement: 1970 to 1984*. Oxford: Pergamon.
- Keeves, J. P. & Bourke, S. F. (1977). *Australian Studies in School Performance Volume II: Literacy and Numeracy in Australian Schools, Final Report*. Canberra: Australian Government Publishing Service.
- Kozma, R. (Ed.). (2003). *Technology, innovation, and education change: A global perspective: A report of the Second Information Technology in Education Study (SITES) Module 2*. Eugene, OR: International Society for Technology in Education.
- Krathwohl, D. R., Bloom, B. S., & Masia, B. B. (1964). *Taxonomy of Educational Objectives. The Classification of Educational Goals. Handbook 2. Affective Domain*. New York: McKay.
- Larkin, A. I., & Keeves, J. P. (1984). *The Class Size Question: A study at different levels of analysis*. Hawthorn, VIC: ACER.
- Law, N., Pelgrum, W., & Plomp, T. (2008). *Pedagogy and ICT Use in Schools around the World: Findings from the IEA SITES 2006 Study*. Hong Kong: Comparative Education Research Centre, University of Hong Kong.
- Lemmens, M., Druker, S. L., Garnier, H. E., Chen, C. & Roth, K. J. (Eds.). (2006). *TIMSS 1999 Video Study Technical Report, Volume 2: Science Study (NCES 2006-015)*, US Department of Education. Washington, DC: National Center for Education Statistics.
- Lindsey, J. K. (1974). A Re-analysis of Class-size and Achievement in the IEA Mathematics Study. *Comparative Education Review*, 18(2), 314-326.
- Lokan, J., Ford, P., & Greenwood, L. (1996). *Maths and Science on the Line: Australian Junior Secondary Students' Performance in the Third International Mathematics and Science Study*. (TIMSS Australia Monograph no. 1). Melbourne, VIC: ACER.

- Lokan, J., Ford, P., & Greenwood, L. (1997). *Maths and Science on the Line: Australian Middle Primary Students' Performance in the Third International Mathematics and Science Study*. (TIMSS Australia Monograph no. 2). Melbourne, VIC: ACER.
- Lokan, J., Ford, P., & Greenwood, L. (1999). *Maths and Science on the Line: Australian Senior Secondary Students' Performance in the Third International Mathematics and Science Study*. (TIMSS Australia Monograph no. 3). Melbourne, VIC: ACER.
- Lokan, J., Hollingsworth, H., & Hackling, M. (2006). *Teaching Science in Australia: Results from the TIMSS 1999 Video Study*. (TIMSS Australia Monograph no. 8). Melbourne, VIC: ACER.
- Masters, G. N. (1997). Partial Credit Model. In J. P. Keeves (Ed.), *Educational Research, Methodology and Measurement: An International Handbook* (2nd edition), (pp. 857-863). Oxford: Pergamon.
- Masters, G., Lokan, J., Doig, B., Khoo, S. T., Lindsey, J., Robinson, L. & Zammit, S. (1990). *Profiles of Learning: The Basic Skills Testing Program in New South Wales*, 1989. Hawthorn, VIC: ACER.
- Masters, G. N., & Forster, M. (1997a). *Literacy Standards in Australia*. Commonwealth of Australia: Department of Employment, Education, Training and Youth Affairs.
- Masters, G. N., & Forster, M. (1997b). *Mapping Literacy Achievement: Results of the 1996 National School English Literacy Survey*. Commonwealth of Australia: Department of Employment, Education, Training and Youth Affairs.
- Ministerial Council for Education, Early Childhood Development and Youth Affairs (MCEECDYA) (2008). *Melbourne Declaration on Educational Goals for Young Australians*. Carlton, VIC: Curriculum Corporation.
- Ministerial Council on Education, Employment, Training and Youth Affairs (MCEETYA) (1999). *National Goals for Schooling in the Twenty First Century*. Melbourne: Curriculum Corporation. [www.curriculum.edu.au/mceetya/adeldec.htm](http://www.curriculum.edu.au/mceetya/adeldec.htm)
- Ministerial Council for Education, Early Childhood Development and

- Youth Affairs (MCEECDYA) (2010). *National Assessment Program: ICT Literacy in Years 6 and 10 (2008)*. Carlton, VIC: Curriculum Corporation.
- Morgan, G. (1982). The Use of the Rasch Latent Trait Measurement Model in the Equating of Scholastic Aptitude Tests. In D. Spearritt (Ed.), *The Improvement of Measurement in Education and Psychology*, pp. 189-212. Hawthorn, VIC: ACER.
- Moss, J. D. (1982). *Towards Equality: Progress by Girls in Mathematics in Australian Secondary Schools*. Hawthorn, VIC: ACER.
- Peaker, G. F. (1967). The Regression Analyses of the National Survey. In Central Advisory Council for Education (England). *Children and their Primary Schools* (Plowden Report), vol 2, appendix 4, pp. 179-221. London: HMSO.
- Peaker, G. F. (1975). *An Empirical Study of Education in Twenty-One Countries. A Technical Report*. Stockholm: Almqvist and Wiksell.
- Postlethwaite, T. N. (1967). *School Organization and Student Achievement. Stockholm Studies in Educational Psychology*. Stockholm: Almqvist and Wiksell.
- Postlethwaite, T. N., & Wiley, D. E. (1992). *The IEA Study of Science II, Science Achievement in Twenty-Three Countries*. Oxford: Pergamon.
- Radford, W. C. (1966). *Staying Longer at School* (ACER Information Bulletin No.47). Hawthorn, VIC: ACER.
- Rosier, M. J. (1978). *Early School Leavers in Australia*. IEA Monograph Studies No.7. Stockholm: Almqvist and Wiksell.
- Rosier, M. J. (1980). *Changes in Secondary School Mathematics in Australia 1964-1978*. Hawthorn, VIC: ACER.
- Rosier, M. J., & Banks, D. K. (1990). *The Scientific Literacy of Australian Students*. Hawthorn, VIC: ACER.
- Rosier, M. J., & Keeves, J. P. (1991). *The IEA Study of Science I: Science Education and Curricula in Twenty-Three Countries*. Oxford: Pergamon.
- Rosier, M. J., & Long, M. G. (1991). *The Scientific Achievement of Year 12 Students in Australia*. Hawthorn, VIC: ACER.

- Ross, K. N. (1983). *Social Area Indicators of Educational Need*. Hawthorn, VIC: ACER.
- Spearritt, D. (Ed.). (1982). *The Improvement of Measurement in Education and Psychology*. Hawthorn, VIC: ACER.
- Stigler, J. W., & Hiebert, J. (1999). *The Teaching Gap: Best Ideas from the World's Teachers for Improving Education in the Classroom*. New York: The Free Press.
- Super, D. E. (1967). *Towards a Cross-national Model of Educational Achievement in a National Economy*. New York: Teachers College, Columbia University.
- Thomson, S., & Fleming, N. (2004a). *Examining the evidence: Science achievement in Australian schools in TIMSS 2002* (TIMSS Australia Monograph no. 7). Camberwell, VIC: ACER.
- Thomson, S., & Fleming, N. (2004b). *Summing it up: Mathematics achievement in Australian schools in TIMSS 2002* (TIMSS Australia Monograph no. 6). Camberwell, VIC: ACER.
- Thomson, S., McKelvie, P., & Murnane, H. (2006). *Achievement of Australia's Early Secondary Indigenous Students: Findings from TIMSS 2003* (TIMSS Australia Monograph no. 10). Camberwell, VIC: ACER.
- Thomson, S., Wernert, N., Underwood, C., & Nicholas, M. (2008). *TIMSS 2007: Taking a closer look at mathematics and science in Australia* (TIMSS Australia Monograph no. 11). Camberwell, VIC: ACER.
- Tyler, R. (1949). *Basic Principles of Curriculum and Instruction*. Chicago: University of Chicago Press. (republished, 1971).
- Walker, D. A. (1967). An attempt to construct a model of the effects of selection. In T. Husén (1967), *International Study of Achievement in Mathematics: Vol 2*, (pp. 135-139). Stockholm: Almqvist and Wiksell.
- Wilson, M. (1989). Saltus: A Psychometric Model of Discontinuity in Cognitive Development. *Psychological Bulletin*, 105 (2), 276-289.
- Wilson, M. (1992). The Ordered Partition Model: An Extension of the Partial Credit Model. *Applied Psychological Measurement*, 16(4), 309-325.

## CHAPTER 15

# Chilean Participation in IEA Studies

*Leonor Cariola, Catalina Covacevich,  
Johanna Gubler, Ema Lagos & Marcela Ortiz*

### Introduction

The first time Chile participated in a standardized external assessment of students' learning at a national level was the IEA Six Subject Study in the 1960s. This may be considered as the first phase of our country's involvement in international evaluations.

In this article, we will review the background of Chilean educational assessments, showing that the number of international studies systematically increased to eventually form part of the national assessment system. Next, we will recount the record of our participation in IEA studies and, finally, we will describe the learning outcomes and uses we have drawn from these experiences.

### International Assessments as Part of an Evaluation System with Certain Tradition

Chile has a long tradition in student assessment and our participation and experience with IEA is linked to this tradition.

We participated in the 'Six Subject Study' (SSS) in the 1960s. Back then, the Statistical Research Institute of the University of Chile, chaired by Erika Grassau, was widely known for its work in the area of educational assessment, mainly thanks to the design of a university entrance test. Following the Institute's dissemination of the results of this test, the IEA invited Erika and her colleagues to participate in the Six Subject Study. Chile accepted the offer and joined in 1967, thus permitting the Institute to become an IEA member.

In 1970, after the field trial and prior to administration of the final test (SSS), the Institute published its experience drawing attention to the value of the study (Grassau & Rodriguez, 1970). The international results were released in 1973, a year of political unrest and change in Chile. This context, together with the fact that the Statistical Research Institute was reorganized, was the reason for the scant impact of the results in Chile. Even though the national results were published in 1975, the publication was only known in academic circles and had little or no effect on public policies (Rodriguez & Menke, 1975).

Another effort for the country's involvement in external evaluations at that time was a government initiative aimed at conducting standardised evaluations at a municipal level. The first students' learning national assessments took place between 1983 and 1985. The Ministry of Education asked the Catholic University of Chile for these evaluations, and subsequently the university became responsible for their development. There was an interval of a few years when no assessments were carried out, but in 1988 the program was re-launched. Since then, SIMCE [which is the National Assessment System in the Ministry of Education] has administered yearly tests to students in the 4th, 8th or 10th grade, and has increased the number of evaluated subjects as well as the frequency of testing. SIMCE was created to gather information through census tests in all schools as input for efforts to improve student achievement at school and system levels.

It was mere coincidence that the bulk of Chile's evaluation efforts came in the 1990s, in the midst of similar attempts undertaken across Latin America, given that the 'evaluation culture' had existed in the country since the early eighties. Chile was certain that assessments could help improve education, that educational outcomes were crucial for economic and social development and that the country should be able to participate competitively in the global economy.

In 1990, a democratic government took office. In the meantime, the coverage of the educational system had grown, but its quality was

questioned. The Ministry of Education was interested in comparing the Chilean curriculum and outcomes with other countries in the world. Hence, at the end of the 1990s, the Ministry was very eager to participate in international assessments; SIMCE, in particular, was keen to participate as it saw an opportunity to complement its results and to improve its own processes.

At the beginning of the 1990s, the IEA was perceived as an important group of professionals dedicated to educational research and evaluation. The country formally re-joined the IEA at the end of the decade, when it took part in TIMSS 1999. Chile's participation was through the Curriculum and Evaluation Unit of the Ministry of Education. It was not a random choice, and it could have been outsourced as had been done with other international studies. But on that occasion, it was thought important to carry out the study within the Ministry as part of a learning process and to obtain first-hand knowledge from the experience. For the first time, the Ministry of Education was in charge of an international assessment at student level.

The above occurred within the context of deep concern with globalization, of looking out to the world and aspiring to international standards and procedures to promote and guide the country's development.

Chile's interest in participating in TIMSS stemmed from a chance meeting at an Asia-Pacific Cooperation Forum (APEC) in Manila in 1997. Cristian Cox, the Curriculum and Evaluation Unit Coordinator at the time, met Hans Wagemaker, IEA Executive Director, and informed José Pablo Arellano, the Minister of Education at that time, of his intention to take part in TIMSS. The proposal received the Minister's unwavering support. After the meeting, María Inés Alvarez, a psychologist and a maths specialist, was hired to coordinate the study and became part of the national assessment team. Later, María Inés Alvarez became the coordinator of LLECE (a Latin-American study conducted by OREALC/UNESCO) and the National SIMCE Coordinator.



Soon after, Cox learnt that IEA was conducting another study at that time: Civic education (CIVED) under the leadership of Judith Torney-Purta. Given the importance of that subject for the historical moment Chile was living at the time, our country also became involved. To coordinate our participation in CIVED, the Curriculum and Assessment Unit hired a sociologist, Leonor Cariola, who shortly after called for the assistance of Ema Lagos. From that time, there has been a special team dedicated to international studies within the Curriculum and Assessment Unit.

SIMCE is now in charge of national and international assessments, both of which are considered part of the national assessment system. Currently, SIMCE coordinates and administers throughout the country several international educational assessment studies, namely, those organized by IEA, OECD and UNESCO.

## **Chile's Participation in IEA Studies**

### ***Six Subject Survey (1966 -1973)***

The subjects covered in the study were Science, Reading Comprehension, Literature, English as a Foreign Language, French as a Foreign Language, and Civic Education. Chile participated in all areas apart from civic education. The target population was 14-year-old students and students in the final grade of secondary school.

### ***The Written Composition Study (1983-1988)***

The study examined teaching and learning of written composition. The study included three populations: students near the end of primary schooling, students near the end of compulsory schooling, and students near the end of academic secondary school.

### ***TIMSS 1999, 2003, 2011***

TIMSS-R (i.e. TIMSS repeat, later called TIMSS 1999) was administered

in Chile in 1998 to a national sample of 7th and 8th graders. It was a source of learning and multiple challenges. A university was asked to coordinate several stages, from the test administration in schools up to the final database filing. TIMSS-R was the first opportunity to observe Chilean students' achievement in maths and science in the light of such a large international comparison. The performance of Chilean students in math and science was well below the performance level of most countries. TIMSS-R was launched within the Ministry rather privately, with a few academics attending and without press coverage.

The second experience with TIMSS took place in 2003. By that time, the International Studies Team had acquired some experience. Hence, the test administration and all the processes were much easier and went ahead smoothly. The last National Research Coordinators (NRC) Meeting for TIMSS 2003 was held in Santiago. More than a hundred people gathered in Chile to review the international report and it was a privilege to make such a prestigious gathering possible.

A few months after the international release, the SIMCE International Studies Team published a national report which included TIMSS 2003 results. In general, the performance levels of Chilean students remained unchanged from the 1999 results in both areas.

Chile did not take part in TIMSS 2007, as at that time the country was participating in PISA 2006 (after skipping PISA 2003) and in an educational research project for Latin American countries called 'SERCE', organized by UNESCO, and the Ministry decided to put resources into data analysis instead of participating in another study after considering that the resources were insufficient to implement and analyse so many studies so frequently.

At present Chile is back to TIMSS 2011, testing 8th grades and, for the first time, 4th grades as well.

### ***CIVED (1998 – 2002)***

Chile's participation in the Civic Education Study included the stan-

dard population of 8th graders and the optional population of 12th graders. Chilean students obtained an average score lower than their international counterparts in the areas of knowledge. The study also assessed students' attitudes towards politics, democracy and citizenship, among other things. In comparison to the international average, Chilean students showed more interest in politics, and were more likely to vote and participate in the activities of the community. However, they also had lower expectations in relation to their participation in conventional political activities, and placed less trust in political parties, courts of justice and municipal governments. The results were scarcely known through the media, and it was not until a year after the release of these results that we held a Seminar with the participation of local academics and published our national report.

### ***ICCS 2009***

The International Civics and Citizenship Education Study was administered in Chile in 2008 to a national sample of 8th graders. Although Chilean students obtained higher scores than their peers in the other participating Latin American countries, their average score was below the international average. There were no improvements in Chilean students' average scores between the 2008 CIVED study and ICCS 2009. In comparison to the international average, Chilean students placed more importance on the role of citizenship as a social movement, showed higher expectations in relation to their participation in schools and in illegal protests in the future and declared positive attitudes towards the rights of ethnic and immigrant groups.

Chile has also participated in other IEA studies that are not directly related to student achievement and that are managed by other departments of the Ministry of Education. We participated in the Second Information Technology in Education Study, Module 2 (SITES-M2), in SITES 2006 and in the Teacher Education and Development Study in Mathematics (TEDS-M).

## Participation Criteria in International Studies

Chile's participation in international studies has been conceived as a complement to the evidence and information generated by the national system of evaluating learning outcomes. The results of the international studies enable the Ministry of Education to place the learning outcomes of the Chilean educational system (in some areas and grades) within an international context. At the same time, they also permit Chile to put the country's learning goals into an international perspective. In this sense, the participation in these studies reinforces the country's aspirations and learning aims setting up new and higher educational standards.

Moreover, the subjects that international studies choose to assess affect public opinion and policies, as these subjects are perceived as relevant by international experts. The IEA stands out among the institutions dedicated to assessment, and the fact that it assesses not only traditional 'academic' subjects, such as reading, science and maths (undoubtedly all highly relevant) but also places importance on civic education, is a strong message aimed at educational policy makers. We see it as an invaluable contribution to curricula, policies and public opinion.

International studies also complement the national system as some will test areas or examine aspects different from those considered in the national evaluation system, for instance, students' attitudes, or other contextual indicators to measure associated factors. This enables us to identify certain special features of our system, for example, Chilean teachers' confidence and assurance when teaching their subjects, and the higher average age of Chilean teachers as indicated in the TIMSS 2003 study.

Additionally, the fact that international studies are carried out on a sample basis gives a broader content coverage of the assessed subject areas, monitoring certain aspects that do not necessarily require universal coverage. Because international study data are not revealed publicly on a school-by-school basis and do not involve consequences

for the participating schools, they can include different aspects without requiring the same statistical reliability at a course or school level.

Evaluations have become increasingly important in the educational context. In Chile, the results of national and international assessments have been used progressively more as evidence to support and report on the development of public policies in education. For example, the preferential subsidy, which by law must be provided to schools for each underprivileged student enrolled, so long as they sign an agreement to improve school outcomes according to the national evaluations.

Although standard assessments cannot evaluate all that is considered important in education, they are a useful tool that contributes to the protection of equity in learning outcomes within countries. At an international level, the above-mentioned studies provide evidence and guidelines on students' education and, looking ahead, on their preparation in relation to job access and incomes in a globalized economic and social context. In this sense, the IEA studies have assessed learning outcomes traditionally related to the productive areas such as maths and science and have also assessed learning outcomes related to active participation in society such as attitudes towards and knowledge of citizenship issues.

### **Organizing our Work**

Since Chile's re-entry into the IEA, the coordination of international studies has been located in the Curriculum and Evaluation Unit of the Ministry of Education, within SIMCE. It was designed in such a way so as to permit the National Assessment System to feed from international studies and permanently receive up-to-date information. After the innovative start to the evaluation policy in the 1980s, the system stagnated, and by the late 1990s it had become outdated. Chile's participation in international studies has led to improvements in areas such as item construction, software development and use, design and coding of open-ended questions, fieldwork procedures, and quality

control among others. The current organization seeks this feedback through the involvement of each SIMCE team.

The studies in which Chile participates are coordinated by a team dedicated to international studies. There is a technical coordinator in charge of the team who facilitates exchanges with other teams, encourages mutual assistance among study coordinators and promotes the analysis of available data. Each study is led by a national research coordinator responsible for managing and overseeing the overall study, including externalized processes. SIMCE has diverse technical teams that support each study whenever required, on issues such as database management, disciplinary approaches, data analysis, etc. Other activities, such as translations, printing and tests administration, are externalised.

### **Difficulties to Overcome and Challenges**

At the beginning, the adherence to international standardized processes posed a challenge: it was necessary to learn new techniques, new methodologies and research skills and to follow requirements to register each step of every process. Fortunately, our international counterparts have always been extremely professional and supportive of the national centres and teams, allowing us to overcome these difficulties without major effort. It is also clear that communication and procedures in international studies have systematically improved after each study, thus making the situation much easier than what it was 10 years ago.

There is an ongoing difficulty related to the complexity of our administrative procedures. For example, budget arrangements must be planned one year in advance, so hiring an institution to administer tests implies we must have detailed information regarding the test instruments considerably in advance. For this reason, we have learned to estimate the resources we may need for now and for at least two years down the road. The challenge that these international studies pose for Chile is related

to the strategic problem of how to transform a tool that is used at a macro-system level into a tool that is useful for teachers and schools. We must find a way to use the results compiled by international studies where they matter most – in the classroom. It can also be a challenge to reveal certain analyses and conclusions that could benefit educational policies, schools and teachers. This is the case when results are ‘bad’, and the Ministry has had to learn how to deal with communicating these results. There are certain political costs that accompany the benefits of knowing more and having better information about students’ learning levels. Nevertheless, Chile seems to have overcome this difficulty and continues to participate in studies and to report the results. In recent years our society has embraced the culture of evaluation and has been able to focus on what needs to be done to improve our education system.

Another current challenge is related to the use of available data and in-depth research. Since the Ministry is unable to undertake all the data analysis resulting from the studies, universities and investigation centres can help by increasing their research related to international studies. As stated earlier, one of the main reasons for participating in these studies is to obtain relevant and supporting evidence for the policy-making processes, at national, school and class levels. Very frequently the information produced by international reports is too general for specific countries. Therefore, national reports can and should intensify the analysis of their country-specific interests. The Ministry of Education has been meeting this challenge progressing in the development of national reports produced by the International Studies Team at SIMCE and by increasingly inviting universities and research centres to use and further analyze the national data available from international studies. An example of this is the book published by the Ministry of Education in 2009 with further analysis from PISA 2006 (Ministry of Education of Chile, 2009). The work done regarding these national reports may be reviewed at the International Studies section of the SIMCE website (<http://www.simce.cl>).

## **Learning from these Experiences**

The most important things that we gained from participation in international studies— beyond access to international reliable information on the educational system and its achievements – are related to learning about curriculum development, methodology for the national assessment system, and professional development for the people involved. Below, we explain in more detail what has been achieved through the IEA studies.

### *Learning for the Curriculum*

At the beginning of the 1990s, Chile implemented a curricular reform followed by an ongoing supervision of its implementation and evaluation of how well it was adapted to the country's needs. Since then, the primary and secondary education curriculum has been updated several times. The updates implemented after 2000 have taken into consideration the content and approaches of the international studies.

It became evident from TIMSS 1999 that our math and science curriculum suffered from an imbalance and, moreover, in the case of maths, the training of local teachers was badly wanting. An analysis of the TIMSS 2003 test and Chilean curriculum showed that despite certain improvements in the reformulated curriculum implemented in 8th grade as of 2002, there was still an important distance between the test and the curriculum, especially in algebra and geometry in maths, and physics and environmental science in science. The more recent (2009) changes to the science and math curriculum have been made with consideration of the content and approaches outlined in the TIMSS study, among other sources.

In 2003, the Ministry of Education, through the Curriculum and Assessment Unit, initiated a process aimed at creating progress maps (content standards) for pre-primary, primary and secondary education. These progress maps were designed to visualize learning progress in different areas of the national curriculum and to describe



the development of the key competencies fostered by the curriculum from 1st through 12th grade. TIMSS and other international studies were used as an input and inspiration for the development of these progress maps.

Soon after creation of the progress maps, benchmarks for achievement were developed, beginning with the 4th grade. These were established with the aim of providing teachers, schools and parents with descriptive information of what students knew and were able to do in the subjects assessed by the SIMCE national tests. As with progress maps, TIMSS international benchmarks were useful as a guide for formulating achievement levels for particular content areas. Among other sources, our student's results on some TIMSS items were used as evidence to identify which tasks required more in-depth knowledge and proficiency for the development of the 8th grade achievement levels.

The nineties curricular reform included a new approach to civic education, which used to be a high school subject, but was now considered a holistic school experience to be incorporated across the curriculum. The way that the IEA study CIVED looked at civic education was a valuable tool to support this new curricular approach and helped us to better understand this new concept in a way that emphasized abilities and attitudes besides knowledge, and considered aspects such as classroom and school context, as well as teaching practices. The Chilean curriculum includes a compulsory subject called 'class assembly'. CIVED's theoretical framework was taken into account when designing the study plans for this subject between 1999 and 2001 and also in the subject's 2002 curricular adjustment. Additionally, in 2004, a commission was set up to examine the way civic education was understood and taught, to propose new requirements and to suggest how to monitor and assess civic education. A document which summarised the vision of the reform on civic education was presented to the commission to be used as the grounds for their work and was strongly influenced by the CIVED framework.

There have also been initiatives to develop special material for teach-

ers. Studies can become guidelines for teacher practices, provided they are thoroughly analysed and the proposals worked out. To support this idea, the Curriculum and Evaluation Unit produced a national document with guiding principles for the subject 'class assembly' where some conclusions from the CIVED study were summarised and analysed. Furthermore, members of the team of the Curriculum and Evaluation Unit considered, among other sources, the framework presented by TIMSS to develop a document to explain curricular aspects to educators (Montecinos, Ortiz, Gómez, Pino, Pino, & Muñoz, 2005). The International Studies Team has made efforts to address teacher needs and facilitate their use of international studies through the publication of national reports which, for example, describe the reference framework of the test and examples of questions that might help teachers in the classroom.

### *Learning for SIMCE*

Chile's renewed participation in international studies led to important improvements in many areas of assessment and evaluation. We learned much about evaluation techniques as well as statistical analysis and operational procedures. For example, the use of the Item Response Theory (IRT) in Chile advanced quickly, thanks to the country's participation in TIMSS 1999. Similarly, international studies also gave us valuable knowledge on item construction, the incorporation of open-ended response questions in national tests, how to develop better coding guides and better ways to code open-ended response questions. From the study manuals we also learned more about standardized methods and procedures for administering the tests, including material organization, general instructions for students, several forms to check and register information, etc. As a result, SIMCE improved testing procedures, including quality control and monitoring in several stages.

SIMCE's technical teams participate in international studies according to their own area of expertise (data management, data analysis, item

construction, etc.) and this has helped SIMCE examine the way things are accomplished and see different ways of doing things. For example, SIMCE's new software for data validation was inspired by the IEA software that included data reliability checks at different stages and levels. SIMCE has also improved context questionnaires by taking into account international examples and they now follow a reviewed and widely discussed framework.

### *Other Impacts*

For all the researchers and professionals involved, participation in international studies has been an enriching experience. Most of these professionals have continued their careers in educational assessment in the Ministry of Education or in other national and international organizations, and most are today highly regarded national and international experts. This has permitted them to retain their knowledge and deepen it within the country and beyond.

Participation in international studies has also led to the creation of several networks and links. It has strengthened the assessment areas in the participating countries by permitting them; for example, to take part in training sessions for specific groups and needs. One example of this is the training session on national reports offered for the Latin American group in the context of the International Civic and Citizenship Study (ICCS) 2009. Furthermore, the creation of a Latin American module in ICCS 2009 had a great impact on establishing regional networks and curriculum comparisons that otherwise would not have happened. The participation of six Latin American countries can also be seen as a consequence of regional cooperation, as it was located in the context of SREDECC, a regional initiative to promote civic education.

This collaboration has also facilitated the implementation of consultancy services and support among countries for example, in ICCS 2009 we received and shared some translated documents (manuals and framework) between different Spanish speaking countries.

### *And When All is Said and Done ...*

We would like to thank the IEA for all the work carried out for so many years, and for the opportunity to learn from them and work together with them. We have met many interesting people with whom we have shared many hours in crowded and over-crowded meetings, with or without luggage...and have especially appreciated the chance to visit and share with people from so many different places of the world.

### References

- Grassau, E. & Rodríguez, C. (1970). *Investigación Internacional sobre la Evaluación del Rendimiento Escolar. [International Research on School Achievement Assessment]*. Santiago: Instituto de Investigaciones Estadísticas, Universidad de Chile.
- Rodríguez, C. & Menke, L. (1975). *Oportunidad de Aprender. [An Opportunity to Learn]*. Santiago: Instituto de Investigaciones Estadísticas, Universidad de Chile.
- Ministry of Education of Chile. (2009). *What does PISA tell us about the education of high school Students in Chile? New analyses and perspectives on the PISA 2006 results*. Santiago: Author. Retrieved from [http://www.simce.cl/fileadmin/Documentos\\_y\\_archivos\\_SIMCE/PISA2006/PISA\\_AinglesLIBRO.pdf](http://www.simce.cl/fileadmin/Documentos_y_archivos_SIMCE/PISA2006/PISA_AinglesLIBRO.pdf)
- Montecinos, C., Ortiz, I., Gómez, H., Pino, J., Pino, M., & Muñoz, G. (2005, Noviembre). *Un currículo para la formación de profesores de kinder a 8° básico que enseñan ciencias. Trabajo presentado en el XVIII Encuentro Nacional y IV Encuentro Internacional de Investigadores en Educación, CPEIP* [A curriculum for the training of pre-primary through 8th grade science teachers. Paper presented at the XVIII National Meeting and IV International Meeting of Educational Researchers, CPEIP]. Santiago.



## CHAPTER 16

# The Significance of IEA Studies for Education in East Asia

*Frederick K.S. Leung*

### Introduction

Since the mid 1990s, education systems in the East Asian region including Chinese Taipei, Hong Kong, Japan, Korea, and Singapore have begun to draw the attention of the education community and policy makers worldwide because of the superior performance of their students in IEA studies (Beaton et al., 1997a, b; Martin et al., 1997, 2000, 2004, 2008; Mullis et al., 1997, 2000, 2004, 2008). Prior to this, these systems (with the possible exception of Japan) were not well known for their education, and they were not of particular interest to the international education community. Given the changed situation, how do we locate the East Asian systems in terms of the stage of their development in education? This is a particularly important question for educators in East Asia as they are participating more and more in the international education arena and may feel the need to seek an identity in education (Leung, 2001). In this paper, East Asia is discussed as a cultural rather than geographic demarcation. East Asian systems are those which share roughly the same culture, referred to as the Confucian heritage culture (CHC) by Biggs (1996).

Notwithstanding the superior performance of East Asian students mentioned above, past literature has indicated that instructional practices in these East Asian systems were rather traditional and backward, failing to keep pace with the latest developments in learning and instructional theories. For example, in the Hong Kong report of the IEA Second International Mathematics Study (SIMS), Brimer and Griffin (1985) made the following observation when commenting on mathematics teaching in Hong Kong in the 1980s:

The schools in Hong Kong have a reputation for reliance on teacher dominated instructional strategies. Pressures arise from the external examinations, the complexities of languages, expansion of the school system ... There has developed a tendency for classes to be taught by lecture-style delivery, with little student participation, apart from note taking and completing assigned work. ... The intense pressure of examinations, the expectations of parents, pupils and colleagues appear to encourage teachers to impart knowledge and instill in students a need to learn for a predominantly recall mode of performance. (Brimer & Griffin, 1985, p.23)

This is consistent with the description of mathematics teaching found in the East Asian systems in general. In the literature, it has been reported that

the curricula in these East Asian countries are content oriented and examination driven. Teaching is very traditional and old fashioned. Teachers in these countries seem to be ignorant about the latest methods of teaching, and think that competence in mathematics alone is sufficient for an effective teaching of the subject. Classroom teaching is conducted in a whole class setting, and given the large class size involved, there are virtually no group work or activities. Instruction is teacher dominated, and student involvement is minimal. Memorization of mathematical facts is stressed and students learn mainly by rote. There is ample amount of practice of mathematical skills, mostly without thorough understanding. Students and teachers are subjected to excessive pressure from the highly competitive examinations, and students do not seem to enjoy their study. (Brimer & Griffin, 1985; Biggs, 1994; Leung, 1995, 2000; Wong & Cheung, 1997; Wong, 1998 - quoted in Leung, 2001, pp.35-36)

This description does not sit comfortably with the superior performance of East Asian students mentioned above, hence the necessity for East Asian educators, and maybe for educators beyond East Asia as well, to confront the issue of an East Asian identity in education (Leung, 2001). And this is where IEA studies may have something to

offer these East Asian systems.

In the remainder of this chapter, the reliability and validity of IEA studies will be discussed, as will the achievements and attitudes of East Asian students in some IEA studies. I will also discuss mathematics teaching in the East Asian classroom as this was revealed in the TIMSS 1999 Video Study. The reasons for the apparent contradiction between teaching and achievement will be discussed, and implications of IEA studies for the East Asian systems will be explored.

### **Reliability and Validity of IEA Studies**

An issue of concern to educators and policy makers in East Asian systems is the question of how well, or how badly, their systems perform in terms of student achievement. With the increasing demand for accountability from the community, most countries worldwide have measures in place to monitor the effectiveness of their own education systems. In addition to whatever national instruments are used, many countries feel the need for an 'objective', accurate and meaningful measure of student achievement to answer the question of effectiveness. The measure must have international credibility as well as national relevance. What is needed is an international study with endorsement from a large number of countries, and this is clearly what IEA studies can provide. Most IEA studies are large-scale ones with many participating countries, but are IEA studies accurate and objective measures of achievement across countries?

IEA has a long tradition of engaging in rigorous and quantitative comparative studies. The reliability of various IEA studies has been established or defended in the technical reports of the corresponding studies. But are IEA studies, in fact, valid?

With this question, we are talking about two levels of one aspect of validity. At the international level, the measure should reflect the latest understanding of the subject matter. If it is a study of science



achievement, for example, then the measure should be based on the latest accepted theories in science education. At the national level, it means that educators in a certain system have to be convinced that the measure coincides with the national understanding of the subject matter. For the former, it is the practice of IEA to involve international experts in all their studies, and hence this level of validity is taken care of to a certain extent. In the attempt to ensure consistency between the measure selected and national understanding, IEA practice has been to rely on national participation (through National Research Coordinators [NRC] meetings, for example). Thus, all systems/countries participating in a given study would be involved in making the major decisions on certain aspects of that study, including the assessment framework and instruments, so that the measure is 'equally fair or unfair' to all involved. Major decisions on individual IEA studies are made during NRC meetings, while decisions on general policies are made at the annual IEA General Assembly (GA) meetings. When the GA is not in session, decisions are made by a Standing Committee, with members elected from amongst GA members. Special groups for IEA studies in general (for example, Technical Executive Group or TEG, Publications and Editorial Committee or PEC) and for particular studies (for example, the Science and Mathematics Item Review Committee [SMIRC] and Questionnaire Item Review Committee [QIRC] of TIMSS) also comprise individuals selected from a variety of the member countries. The active participation of researchers from different IEA systems is the best guarantee for this second level of validity of the IEA studies.

In addition, in each round of TIMSS, a 'Test-Curriculum Matching Analysis' (TCMA) is performed to investigate "the extent to which the ... assessment was relevant to each country's curriculum" and "the impact on a country's performance of including only achievement items that were judged to be relevant to its own curriculum" (Mullis et al., 2008, p. 440). Taking TIMSS 2007 Mathematics Study as an example, the international report pointed out that the "proportion of items

judged appropriate (by each participating country) was fairly high" (Mullis et al., 2008, p. 441). Understandably, "countries performed better on their own item sets than on the items overall", but the report added that they performed better "not by much" (p. 442). The report concluded that "It is clear that the selection of items does not have a major effect on the relative performance among TIMSS participants", and that "The TCMA results provide evidence that the TIMSS ... assessment provides a reasonable basis for comparing achievement of the participating countries" (p. 443). So by and large, we can say that the goal of devising a measure which is 'equally fair or unfair' to all the participating systems is achieved.

### **Student Achievement in East Asian Systems**

A number of IEA studies have been conducted since the 1990s; in this paper, however, only the following series of studies will be discussed:

TIMSS: 1995, 1999, 2003, 2007

PIRLS: 2001, 2006

TIMSS Video Study: 1995, 1999

How well did East Asian students perform in these studies? A cursory reading of the international reports of the above studies indicates that students from East Asian systems had consistently high performance in mathematics in all four rounds of TIMSS. In the area of science, apart from Hong Kong, their performance was also high in the 1990s, and in the latest two rounds of TIMSS in 2003 and 2007, all East Asian systems performed highly. In the area of reading, the PIRLS results of East Asian systems in 2001 were only moderate, but in 2006, their performance was high for all systems except Chinese Taipei, which joined PIRLS for the first time in 2006. So all in all, we can conclude that the performance of East Asian students in these IEA studies since the 1990s has been extremely impressive (see Table 1).

**Table 1.** Performance of East Asian students in IEA studies

Subject	Years	Performance
Mathematics	1995/1999/2003/2007	All consistently high
Science	1995/1999	High except for Hong Kong
	2003/2007	All high
Reading	2001	Moderate
	2006	High except for Chinese Taipei

It should be pointed out that the superior achievements of East Asian students reported above were somewhat unexpected. The author remembers that, at the first TIMSS NRC meeting in Vancouver in 1991, when the international study coordinator reminded NRCs that TIMSS was essentially not a contest, NRCs joked about competing against the 'strong' countries. They joked about beating the Germans, the Americans, the Hungarians, etc.; but no one mentioned the East Asian systems, except for Japan. So as mentioned at the beginning of this paper, in the eyes of these NRCs at least, the East Asian systems were not known for their education.

### **Attitudes of East Asian Students in TIMSS and PIRLS**

However, the high achievements of East Asian students discussed above do not seem to be accompanied by correspondingly positive attitudes towards study. Taking TIMSS 2007 and PIRLS 2006 as examples, one can summarize the attitudes of East Asian students towards mathematics, science and reading as follows:

- TIMSS 2007: East Asian students (other than those in Singapore) did not value mathematics and science highly, and all of them (including those from Singapore) lacked self-confidence in learning mathematics and science.
- PIRLS 2006: East Asian students' attitudes towards reading were moderately high, but their self-concept in reading was low. (See also Table2.)

**Table 2.** East Asian students' attitudes towards study

Subject	Attitude	Rating
TIMSS 2007 Mathematics	Valuing of Mathematics	Low
	Self-Confidence in Learning Mathematics	Low
TIMSS 2007 Science	Valuing of Science	Low
	Self-Confidence in Learning Science	Low
PIRLS 2006	Attitudes towards Reading	Medium
	Reading Self-Concept	Low

### Mathematics Teaching in East Asian Classrooms

How do we explain the high achievement and yet relatively negative attitudes of East Asian students in the above-mentioned IEA studies? Both TIMSS and PIRLS have a questionnaire component – for the system, school, teacher and students-- which may be used to explore the relation between various background factors on the one hand, and student achievements and attitudes on the other. The TIMSS and PIRLS use a three-level model of the curriculum: intended, implemented and attained curriculum, and the questionnaires measure variables at the intended and implemented levels, plus the attitudes of students at the attained level. Since students obtain most of their academic knowledge in the classroom, one obvious variable at the implemented level is that of classroom teaching, potentially the most important variable in explaining student achievement. It is for this very reason that another series of IEA studies was devoted to a more thorough examination of this variable: the TIMSS Video Study. Three IEA video studies have been conducted so far: a 1995 mathematics study (Stigler et al., 1999), a 1999 mathematics study (Hiebert et al., 2003) and a 1999 science study (Roth et al., 2006). In this paper, we confine our discussion to the 1999 mathematics video study.

#### *TIMSS 1999 Video Study (Mathematics)*

The goal of the TIMSS 1999 Video Study was to “describe and compare

eighth-grade mathematics teaching across seven countries” (Australia, the Czech Republic, Hong Kong SAR, Japan,<sup>1</sup> The Netherlands, Switzerland, the United States). The TIMSS Video Study is unique in the sense that it was the first video study based on a national probability sample of 8th grade mathematics lessons (target sample size was 100 lessons per system) in the participating systems. It is in this sense that it was the first **Video Survey** in the history of educational research.

Altogether the achieved sample in the 1999 mathematics study was 638 videotaped lessons, including the 50 lessons from the 1995 survey in Japan. The data were analyzed both quantitatively, by an international team of researchers from all the participating systems who developed 45 codes and applied them to the video data; and qualitatively, by a specialist group of mathematicians and mathematics educators (known as the mathematics quality analysis group) who reviewed and made judgements on the detailed descriptions of a subset of the lessons ‘country-blind’ – i.e., with all indicators that might reveal the identity of the countries removed.

### **Findings: Instructional practices in East Asia as revealed by analysis of the codes**

From the quantitative analysis of the video data, the instructional practices in the East Asian classroom, as compared to those in the Western classroom, can be characterized by the following (see Leung, 2005):

- Dominance of teacher talk
- Students have more opportunities to learn new content
- Students solve problems that are more complex and are unrelated to real life
- More time spent on mathematical proofs

---

1. The 1995 Japanese data were re-analyzed using the 1999 methodology in some of the analyses.

### **Findings: Quality of the content as judged by the Mathematics Quality Analysis Group**

In comparison with lessons in Western countries, the quality of the content of the Hong Kong lessons (the only East Asian system in this analysis) as judged by the expert group based on a subset of the same data set can be characterized as follows (Leung, 2005):

- Relatively advanced content
- More deductive reasoning
- More coherent
- More fully developed presentation
- Students are more engaged, and
- Overall quality is high

### ***Two different pictures?***

Is the apparent inconsistency between the instructional practices as revealed by the two different analyses of the same video data set due to a conflict between a quantitative versus a qualitative method of data analysis? This touches on methodological issues concerning analysis of video data – it seems that there is an inherent trade-off between reliability and validity when analyzing video data. In any case, the discrepancy of the findings shows that classroom teaching is a complex phenomenon, and a combination of quantitative and qualitative analyses, as adopted in the TIMSS Video Study, may yield findings closer to what really happened in the classroom.

In sum, despite the positive judgement of the Mathematics Quality Analysis Group on the quality of mathematics teaching in the East Asian systems, the quantitative analysis of the TIMSS video data indicates that mathematics teaching in these systems was rather teacher dominated and traditional.

## Discussion

If classroom teaching in East Asia is indeed backward and traditional, how does this backward teaching produce students who perform so well in IEA studies? Why is the superior achievement of East Asian students not accompanied by a correspondingly positive attitude towards study? What are the implications of the findings of these IEA studies for East Asian systems?

The high achievement of East Asian students and the seemingly backward and traditional teaching in the East Asian classroom perhaps remind us that classroom teaching, although an important factor for explaining student achievement, is not the only factor - it may not even be the most important factor. This applies to other background factors as well. To get the most out of IEA studies, factors which explain within-system differences in achievement as well as factors which contribute to across-system differences need to be explored. In examining the findings of IEA studies for the East Asian systems in contrast to other systems, it is obvious that many factors (e.g., SES, attitudes towards the subject matter, etc.) may explain achievement variations within a system, but they do not account for across-system differences. Factors which account for within-system differences are potentially important information for educators and policy makers, and so (and maybe even more so) are factors which account for across-system differences. One important factor which may account for across-system differences, and which is sometimes ignored in discussions on comparative studies of student achievement, is that of the cultural values held by students and teachers. The case for East Asian systems will be discussed below.

### *Some relevant cultural values in the East Asian systems*

#### **Emphasis on Education and High Expectations on Student Achievement**

The first pertinent cultural value in the East Asian systems is the

emphasis placed on education. Lee (1996) pointed out that this emphasis on education rests on the Confucian presumption that everyone is "educable" and "perfectible" (pp.28-30). East Asian parents and teachers attach great importance to the education and achievement of their children and students (Sollenberger, 1968), and there is a high expectation for East Asian students to achieve. This is consistent with the observation by Shen (2005) that schooling or education in these East Asian systems "occupies a far more central position in adolescents' lives than in the U.S." (p.195). It is possible that this emphasis on education and achievement may explain, at least in part, the superior performance of East Asian students in IEA studies.

### **Examination culture**

Perhaps related to the emphasis on education, another relevant cultural value in the East Asian systems is the heavy emphasis placed on examinations. The East Asian systems in this paper are all, to different extents, under the influence of the Chinese culture, and China was the first country in the world to have a system of national examinations. With the Sui Dynasty (A.D. 587), a national examination was instituted in the imperial court to select scholars for high government offices (Ding & Zhang, 1989). From then on,

the examinations at different dynasties were invariably the means to select appointees to the officialdom. ... The examination was later developed into a stratified system where scholars competed in local examinations and became qualified for higher level examinations ... Local successful candidates were awarded lifelong titles of scholars who became local intellectuals with respectable social status. The champions in the examination held at the central imperial court were granted high positions in the government (as high as the prime minister) and often granted marriage to the royal family. (Cheng, 1994)

So through the ages, there has developed in the East Asian culture a great trust in the public examination as a legitimate and fair method of differentiating the successful from the unsuccessful. Has this exami-



nation culture anything to do with the superior performance of East Asian students in IEA studies? Do East Asian students do better in IEA studies simply because they are good in taking examinations, notwithstanding the fact that the IEA tests are low-stake ones?<sup>2</sup> This poses serious challenges to IEA studies, as it touches on another aspect of the validity of these studies: are the IEA studies merely measuring examination skills? Are the instruments of these studies really appropriate for measuring student achievement of the subject matter (e.g., science) across systems?

On the other hand, if tests such as those used in IEA studies are not the appropriate instruments for measuring achievement, what are the alternatives? As pointed out at the beginning of this chapter, IEA studies represent a common agreement among participating countries on how achievement should be measured. If the NRCs for a certain study feel that the assessment framework and study instruments cannot measure what they are supposed to measure, they should suggest alternatives at the NRC meetings until a consensus is reached. Since the assessment framework and instruments of each IEA study are approved by the NRCs concerned, it must be assumed that the participating countries endorse the appropriateness of the instruments. This assumption again points to the importance of member countries' participation in policy making in IEA studies.

### **Practice, memorization, and understanding**

Under the East Asian examination culture, how do the students prepare for examinations? More pertinent to the discussion in this paper, do they prepare themselves (or do their teachers prepare them) for IEA studies the way they prepare for their national examinations?

---

2. The author is the TIMSS NRC for Hong Kong, and his observation is that students in Hong Kong did not take the TIMSS tests very seriously, knowing that they were low-stake tests. However, personal communications with the NRCs in Korea and Singapore indicate that many students in the two countries took the TIMSS tests rather seriously.

A common cultural value held in the East Asian systems is the emphasis on diligence and effort in study (Park & Leung, 2003). There is a greater tendency in East Asian systems than in other systems to attribute students' academic success to effort rather than to innate ability (Leung, 2001), and 'practice makes perfect' is a common dictum in East Asian systems (Li, 2006). Related to this emphasis on diligence is the emphasis placed on practice. Excessive and repetitive practice is sometimes characterized in the West as rote learning. However, in East Asia, "repetitive learning" is conceived as "continuous practice with increasing variation" (Marton, 1997), and practice and repetition are considered a route to understanding (Hess & Azuma, 1991). Memorization, even before thorough understanding, is not conceived as negatively as in other cultures. From the perspective of the East Asian culture, equating memorization without full understanding to rote learning is too simplistic a view. As Leung (2001) pointed out, "understanding is not a yes or no matter, but a continuous process or a continuum. .... Learning is an iterative process of repeated practice, memorizing and understanding" (pp. 40-41). Within the East Asian culture, practice and memorization are accepted as legitimate (and probably effective) means for understanding and learning.

### **Pragmatic philosophy in East Asia**

East Asians are known to be pragmatic people (Shusterman, 2004). Students take a pragmatic approach in their learning, and teachers take a pragmatic approach in their teaching. This may to a certain extent explain the 'traditional teaching' in East Asian classrooms. For example, Leung and Park (2002), in a study which involved interviews with Hong Kong and Korean teachers, reported that "East Asian teachers are competent in mathematics in general, but they deliberately taught in a procedural manner for pedagogical reasons and for the sake of efficiency. Teachers believe that it would be inefficient or confusing for school children to be exposed to rich concepts instead of clear and simple procedures" (p. 126).

As a result, East Asian teachers are very skeptical about high-sounding new theories of teaching and learning, and tend to stick to established practices that have proved to be working. The superior performance of their students merely reinforces such pragmatism and conservatism.

### *The negative attitudes of East Asian students*

Some of the cultural values discussed above, and some other values not yet mentioned (e.g., the virtue of modesty – see the discussion below), may also explain the negative attitudes of students in East Asian systems. These will be discussed below.

#### **Examination culture**

As far as student learning is concerned, the examination culture in East Asian systems is a two-way sword. It may have contributed to the high achievements of East Asian students in international studies, but the examination culture and the consequent competitive examination system may create “undue pressure upon students, resulting in all sorts of harmful effects such as loss of interest in (study) and behavioral problems” (Leung, 2001, p. 43).

Related to this, learning or studying is considered a serious endeavour in the East Asian culture, and students are expected to put in hard work and perseverance (Garvey & Jackson, 1975). This may explain the negative attitudes towards study, as found in various IEA studies.

#### **Intrinsic vs extrinsic motivation**

The examination culture legitimizes performance in examinations as a source of motivation for study. This contrasts with the views of some Western educators who value intrinsic motivation to the extent that they consider extrinsic motivations for study such as those derived from examination pressure as harmful to learning. In contrast, as mentioned above, the Confucian culture emphasizes the importance of education and a high expectation to achieve, and East Asian parents

and teachers may communicate this expectation to their children in an explicit or implicit manner. The high expectations and competitive examinations may have resulted in a large number of students being classified as failures, and repeated experiences of a sense of failure may have reinforced the lack of confidence as found in the IEA studies (Dindyal, 2008; Leung, 2002).

### **The virtue of modesty**

Another virtue highly valued in East Asia is that of modesty, and the justification of this value is reflected in the Chinese dictum 'Contentedness leads to loss; modesty leads to gain'. Children in East Asia are taught from a young age that one should not be boastful. This may have inhibited East Asian students from rating themselves too highly in attitude questionnaires in IEA studies, and the scores may represent less than what the students really think about themselves.

On the other hand, if students are constantly being taught to rate themselves low, they may internalize the idea and end up with really low levels of confidence. In any case, low confidence is not conceived as negatively in East Asia as in other cultures. According to East Asian values, a **negative correlation between students' confidence and achievement is something to be expected: over-confidence lowers the incentive to learn** (Leung, 2002).

### ***Implications of IEA studies for East Asian Systems***

One of the purposes of IEA studies is for systems less knowledgeable in research methods to acquire the expertise of large-scale educational research, and many East Asian systems first took part in IEA studies with such a purpose in mind. In addition, acknowledging that their education systems were 'less developed' than those in Western countries, East Asian systems were also looking to IEA studies to provide 'objective' measures of achievements in order to benchmark their students' performance with the rest of the world, as mentioned at the beginning of this paper. So in accordance with the East Asian attitude

of modesty, many East Asian systems participated in IEA studies with a keen desire to learn, and they do learn much through their participation.

The unexpected superior performance of their students in IEA studies offered these East Asian systems an added bonus to what they learned from these studies. The superior performance assured educators in the East Asian systems that their education was not that bad after all, even when measured against an international standard. This contributed much to the establishment of national pride and confidence in the area of education.

However, the purpose of IEA studies is not simply to bring contentedness to the 'successful' systems; it should instead help educators and policy makers identify factors that contribute to high achievement so that they can devise measures to improve their education systems. An important message arising from the discussions in this paper is that there are different ways to achieve success. East Asian systems are **not inferior (or superior) to, but different from**, Western systems. Their students' success in IEA studies has earned them a place on the international map of education.

The discussions in this paper also alluded to the role of culture in explaining student achievement and educational practices, a factor which is often overlooked by educators and researchers in comparative studies, including IEA studies. Rarely have findings of IEA studies been discussed with respect to cultural differences. The consistently superior performance of East Asian students in IEA studies points to the powerful role culture may play in accounting for across-system differences in educational achievement.

If culture is indeed an important factor, there are a number of implications. Results of IEA studies are meant for educators and policy makers to learn from other systems, but in doing so, the challenge is to ensure that the strengths of one's own system are not compromised in the process. It might be that very complicated cultural factors affect-

ed classroom practices and student achievement, and drastic changes in educational policies and classroom practices should not be undertaken until such factors are thoroughly examined. When adopting practices from other systems that are considered effective, it is important to remember that educational practices cannot be transplanted without regard to the cultural differences. Culture by definition evolves slowly; there is simply no quick transformation of culture. So in making use of the results of IEA studies, one needs to identify not only superficial differences in educational practices, but also the intricate relationships between the practices and the underlying cultural values. By examining these relationships, one may understand better the interactions between educational practices and culture, and it is through identifying commonalities and differences of both educational practices and the underlying cultural values that one may begin to determine how much can or cannot be borrowed from another culture.

For East Asian systems, results of IEA studies provide participating systems the opportunity to evaluate their own cultural values. One should take these values as given, and design and improve educational practices with this in mind. For East Asian systems, the most important lesson they learned from their participation in IEA studies was to identify the factors underlying their strengths, based on the findings of the IEA studies. Theorization on the basis of these identified factors should help educators in the East Asian systems establish their own unique identity in education, instead of relying solely on Western education theories. This is an important step for developing an East Asian pedagogy based on a sound East Asian theory of education, and this is important not only for East Asia, but also for systems in other cultures as well.

## Conclusion

Public interest in IEA studies is usually focused on the relative position of countries in the generated league tables, but IEA studies are not

really competitions in achievement. Results of IEA studies should serve as mirrors for us to better understand our own system. Education is a complex endeavour, and one cannot expect IEA studies to produce quick answers for all national problems in education. What IEA studies can provide are rich data sets with which individual countries can seek answers to problem and issues they are facing, and in this process, **what is needed is wisdom, not just data!**

### References

- Beaton, A. E., Mullis, I. V. S., Martin, M. O., Gonzalez, E. J., Kelly, D. L., & Smith, T.A. (1997a). *Mathematics Achievement in the Middle School Years*. Chestnut Hill, MA: International Study Center, Boston College.
- Beaton, A. E., Martin, M. O., Mullis, I. V. S., Gonzalez, E. J., Smith, T. A., & Kelly, D.L. (1997b). *Science Achievement in the Middle School Years*. Chestnut Hill, MA: International Study Center, Boston College.
- Biggs, J. B. (1994). What are Effective Schools? Lessons from East and West (The Radford Memorial Lecture). *Australian Educational Researcher*, 21, 19-39.
- Biggs, J. B. (1996). Western Misconceptions of the Confucian-Heritage Learning Culture. In D. A. Watkins & J. B. Biggs (Eds.), *The Chinese Learner*. Hong Kong: Comparative Education Research Centre.
- Brimer, A., & Griffin, P. (1985). *Mathematics Achievement in Hong Kong Secondary Schools*. Hong Kong: Centre of Asian Studies, the University of Hong Kong.
- Cheng, K. M. (1994). Quality of Education as Perceived in the Chinese Culture. In T. Takala (Ed.), *Quality of Education in the Context of Culture in Developing Countries*, (pp.67-84). Tampere: Tampere University Press.
- Dindyal, J. (2008). An Overview of the Gender Factor in Mathematics in TIMSS-2003 for the Asia-Pacific Region. *ZDM The International Journal on Mathematics Education*, 40 (6), 993-1005.

- Ding, S. S., & Zhang Z. G. (1989). *Mathematics and Education*. Changsha: Hunan Education Press. [in Chinese]
- Garvey, A., & Jackson, B. (1975). Chinese Children. *Research and Action Project into the Needs of Chinese Children*. Cambridge: National Education Research Development Trust.
- Hess, R., & Azuma, H. (1991). Cultural Support for Schooling: Contrasts between Japan and the United States. *Educational Researcher*, 20 (9), 2-8.
- Hiebert, J., Gallimore, R., Garnier, H., Givvin, K. B., Hollingsworth, H., Jacobs, J., Chui, A. M. Y., Wearne, D., Smith, M., Kersting, N., Manaster, A., Tseng, E., Etterbeek, W., Manaster, C., Gonzales, P. & Stigler, J. (2003). *Teaching Mathematics in Seven Countries. Results From the TIMSS 1999 Video Study*. Washington DC: National Center for Education Statistics.
- Lee, W. O. (1996). The Cultural Context for Chinese Learners: Conceptions of Learning in the Confucian Tradition. In D.A. Watkins & J.B. Biggs, *The Chinese Learner*. Hong Kong: CERC and ACER.
- Leung, F. K. S. (1995). The Mathematics Classroom in Beijing, Hong Kong and London. *Educational Studies in Mathematics*, 29, 297-325.
- Leung, F. K. S. (2000). An Exploration into the Reasons for the High Achievement of East Asian Students, paper presented at the *International Conference on Mathematics Education, History of Mathematics and Cultural History of Mathematics, Informatics and Learning Obstacles*, Beijing, July.
- Leung, F. K. S. (2001). In Search of an East Asian Identity in Mathematics Education. *Educational Studies in Mathematics*, 47, 35-52.
- Leung, F. K. S. (2002). Behind the High Achievement of East Asian Students. *Educational Research and Evaluation*, 8 (1), 87-108.
- Leung, F. K. S. (2005). Some Characteristics of East Asian Mathematics Classrooms Based on Data from the TIMSS 1999 Video Study, *Educational Studies in Mathematics*, 60 (2), 199-215.
- Leung, F. K. S., & Park, K. (2002). Competent Students, Competent



- Teachers? *International Journal of Educational Research*, 37, 113-119.
- Li, S. (2006). Practice Makes Perfect: A Key Belief in China. In F.K.S. Leung, K.D. Graf and F. Lopez-Real (Eds.), *Mathematics Education in Different Cultural Traditions: A Comparative Study of East Asia and the West*, The 13th ICMI Study, New York: Springer, 129-138.
- Martin, M. O., Mullis, I. V. S., Beaton, A. E., Gonzalez, E. J., Smith, T. A. & Kelly, D.L. (1997). *Science Achievement in the Primary School Years*. Chestnut Hill, MA: International Study Center, Boston College.
- Martin, M. O., Mullis, I. V. S., Gonzalez, E. J., Gregory, K. D., Smith T. A., Chrostowski, S. J., Garden, R. A. & O'Connor, K. M. (2000). *TIMSS 1999 International Science Report*. Chestnut Hill, MA: International Study Center, Boston College.
- Martin, M. O., Mullis, I. V. S., Gonzalez, E. J., & Chrostowski, S. J. (2004). *TIMSS 2003 International Science Report*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Martin, M. O., Mullis, I. V. S., & Foy, P. (2008). *TIMSS 2007 International Science Report*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Marton, F. (1997). *Student learning: East and West*. Public lecture presented at The Chinese University of Hong Kong, Shatin, Hong Kong, March.
- Mullis, I. V. S., Martin, M. O., Beaton, A. E., Gonzalez, E. J., Kelly, D. L. & Smith, T.A. (1997). *Mathematics Achievement in the Primary School Years*. Chestnut Hill, MA: International Study Center, Boston College.
- Mullis, I. V. S., Martin, M. O., Gonzalez, E. J., Gregory, K. D., Garden, R. A., O'Connor, K. M., Chrostowski, S. J. & Smith T. A. (2000). *TIMSS 1999 International Mathematics Report*. Chestnut Hill, MA: International Study Center, Boston College.
- Mullis, I. V. S., Martin, M. O., Gonzalez, E. J., & Chrostowski, S. J. (2004). *TIMSS 2003 International Mathematics Report*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Mullis, I. V. S., Martin, M. O. & Foy, P. (2008). *TIMSS 2007 International*

- Mathematics Report*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Park, K. M. & Leung, F. K. S. (2003). Factors Contributing to East Asian Students' High Achievement in Mathematics: The Case of Korea. *The Mathematics Educator*, 1, 7-19.
- Roth, K. J., Druker, S. L., Garnier, H., Lemmens, M., Chen, C., Kawanaka, T., Rasmussen, D., Trubacova, S., Warvi, D., Okamoto, Y., Gonzales, P., Stigler, J., & Gallimore, R. (2006). *Teaching Science in Five Countries: Results From the TIMSS 1999 Video Study*. (NCES 2006-011). U.S. Department of Education. Washington, DC: National Center for Education Statistics.
- Shen, C. (2005). How American Middle Schools Differ from Schools of Five Asian Countries: based on cross-national data from TIMSS 1999. *Educational Research and Evaluation*, 11(2), 179-200.
- Shusterman, R. (2004). Pragmatism and East-Asian Thought. *Metaphilosophy*, 35 (1-2), 13-43.
- Sollenberger, R. T. (1968). Chinese-American child-rearing practices and juvenile delinquency. *Journal of Social Psychology*, 74, 13-23.
- Stigler, J. W., Gonzales, P., Kawanaka, T., Knoll, S., & Serrano, A. (1999). *The TIMSS Videotape Classroom Study: Methods and Findings from an Exploratory Research Project on Eighth-Grade Mathematics Instruction in Germany, Japan, and the United States*. (NCES 1999-074). U.S. Department of Education. Washington, DC: National Center for Education Statistics.
- Wong, N. Y. (1998). In Search of the "CHC" Learner: Smarter, works harder or something more? In H. S. Park, Y. H. Choe, H. Shin, & S. H. Kim (Eds.). *Proceedings of ICMI – East Asia Regional Conference on Mathematical Education*, 1: 85–98. Seoul, Korea: Korean Sub-Commission of ICMI; Korea Society of Mathematical Education; Korea National University of Education.
- Wong, K. M. & Cheung, W. W. (1997). A survey of the Current State of Primary Mathematics Teaching in Hong Kong. *EduMath*, 4, 3-15.



## CHAPTER 17

# The Impact of IEA on Educational Policy Making in Germany

*Rainer Lehmann*

### Introduction

Among the IEA member countries, Germany represents a rather unique case. Straddling the border between 'western democracies' and 'people's republics', during the Cold War it was divided into two rival states. Correspondingly, until 1990, two very different approaches to educational policy making and administration coexisted. Whereas the former West Germany, named Federal Republic of Germany, constituted eleven states each exercising jurisdiction over its own system of education, the East, which had chosen the name German Democratic Republic, had opted for a system under strict centralized control, exercised by the *Sozialistische Einheitspartei Deutschlands* (SED), i.e., the communist party. Unlike Hungary and Poland, the German Democratic Republic (GDR), as long as it existed, never participated in any of the international large-scale assessments. In contrast, the Federal Republic had a history of participation in such studies, beginning with the IEA Pilot Twelve-Country Study (Foshay et al., 1962), and quite a number of studies since.

This context set the stage for the interactions between researchers and policy makers in Germany during the five decades since the emergence of IEA. As will be seen, the findings from the IEA studies did not always generate much interest in Germany, but when they did they were often controversial, especially among politicians who were eager to use - or abuse - them as causal arguments confirming their own preconceptions. The strongly stated assumptions built into the early IEA studies, namely that they could provide quasi-experimental evidence fit to resolve political controversies (Husén, 1967), served per-

haps as an invitation to such a selective, partisan view. It should be noted, however, that the IEA studies also had much deeper indirect effects on the general rationale of educational policy making, very much beyond the mere 'application' of particular findings. A case will be made here for the significance of such impact in Germany. This will be done by distinguishing an early stage of half-hearted and intermittent German participation (1959 – 1988) from a phase of IEA's growing influence (1988 – 1999) and an era of consolidation (from 2000 onwards).

### **The Early Years (1959 – 1988): Half-hearted and Intermittent German Participation**

When in 1958 that remarkable meeting at the UNESCO Institute of Education in Hamburg (now UNESCO Institute for Lifelong Learning) was convened, in which the idea of comparative large-scale assessments was born (see Postlethwaite, 1995), several historical factors clearly favored (West) German participation in such activities, the vicinity of the venue probably being the least important.

First and foremost, the educational discourse of the time was clearly reform oriented. In 1953, the independent German Commission for Education (*Deutscher Ausschuss für das Erziehungs- und Bildungswesen*) was established, which during its existence until 1965 suggested fundamental changes to the structure of the German system of education, most notably – in 1959 – a "Framework for the Reform and Standardisation of the Public School System" (*Rahmenplan zur Umgestaltung und Vereinheitlichung des allgemeinen öffentlichen Schulwesens*). As was the case in other European countries during that period, the respective recommendations wanted legitimacy, preferably in the form of empirical evidence. Instead of referring to current practice and traditions, perceived as inadequate, international comparisons appeared to demonstrate what was possible under reformed conditions. Stated in the words of Foshay et al.: "If custom and law

define what is educationally allowable within a nation, the educational systems beyond one's national boundaries suggest what is educationally possible" (1962, p.7).

A second factor conducive to Germany's involvement in IEA studies is related to a growing awareness of the need for educational reforms. Adjacent academic disciplines - most notably, on the basis of non-technical economic considerations, Georg Picht's proclamation of an "educational catastrophe" in Germany (1964) and, from a sociological perspective, Ralf Dahrendorf's call for education as a "civil right" (1965) - clearly prepared the ground not only for far-reaching political decisions, but also justification for accompanying research. Both lines of later argument - concerns over inadequate achievement levels not meeting the society's needs for qualifications and the emphasis on the equality of educational opportunity - reflect almost exactly the earlier concerns. It will be shown here how IEA has served important functions to substantiate these claims.

A third and final element favouring Germany's early involvement in IEA studies can be seen in the existence of the required research infrastructure. As early as 1951, the German Academy of International Educational Research, located in Frankfurt/Main, had been founded, having some funds from United States sources at its disposal. During the first years of its development, high priority was devoted to the acquisition of state-of-the art technologies of educational measurement. Seen historically, this amounted to regaining access to the international scientific community, an option which had been interrupted by National Socialist rule and terror. Thus, by the end of the 1950s, this agency was technically in a position to enter into the collaborative research efforts of the emerging IEA. Walter Schultze was the National Coordinator for the Twelve-Country Pilot Study and FIMS, the First International Mathematics Study (Husén, 1967). He served until his retirement as the first German Representative in the IEA General Assembly, whereupon his institution - now the German Institute for International Educational Research (*Deutsches Institut für Internationale*

*Pädagogische Forschung*: DIPF) – withdrew from IEA. For more than a decade, no successor in this position was appointed. It was not until 2007 that another Director of DIPF, this time Eckhard Klieme, represented Germany in the General Assembly.

German engagement in the early IEA studies was only intermittent and at best ‘half-hearted’ (Döbrich, 2002; van Ackeren, 2002; Bos, Postlethwaite & Gebauer, 2010). In fact, there are very few traces of original work pertaining to the early IEA studies. Apart from the international volumes reporting on FIMS, only one single German text, produced as an internal project report, can be identified which discusses findings based on this study (Schultze & Riemenschneider, 1967). The sample was restricted to two out of the eleven West German states (Hesse and Schleswig-Holstein) and thus offered virtually no basis for generalisations. Not surprisingly, very little public attention was paid, despite the fact that the limited results suggested that the sampled 7th graders had performed at considerably lower levels than expected. What was left in the educational community, if not in the public mind, was Hans Freudenthal’s critique culminating in the sarcastic acronym IEA = id est absurdum (see Seibel & Wolf-Seibel, 1976).

Even so, Walter Schultze continued to coordinate what might be labelled a ‘Three Subject Survey’ in the set of IEA large-scale assessments known as the Six Subject Survey. While Germany left out Reading Comprehension (Thorndike, 1973), Literature (Purves, 1973) and French as a Foreign Language (Carroll, 1975), only Civic Education (Torney, Oppenheim & Farnen, 1975), Science (Comber & Keeves, 1973) and English as a Foreign Language (Lewis & Massad, 1975) were – partially – implemented in this country. Again, the publication of results was primarily left to the international volumes, the German reporting being limited to internal DIPF documents (science: Schultze, 1974a; English: Schultze, 1975). A single bibliographic trace can be found for the German component to the IEA Civic Education Study: a special analysis published by the Englishman A. N. Oppenheim, one of the principal international investigators, in a

German journal (1976). Generally speaking, only minimal interest is recorded with respect to the German academic community. Seibel & Wolf-Seibel (1976) mention that during the first two years after the release of the data tapes, only two German universities had requested access. One of these exceptions merits further reference below.

The conspicuous absence of any sustained public interest has resulted in – and continues to result in – different explanations. Following early reservations expressed by Schultze & Riemenschneider in their presentation of FIMS results, some have argued that the investments in large-scale international assessments may not have appeared justified in the eyes of policy makers who harboured only vague expectations (van Ackeren 2002). From a historical point of view, Heinz-Elmar Tenorth (2010) has identified a disappointing gap between the promises attached to paradigmatic strands in education and the results produced (this argument not being restricted to empirical research). Also, the traditional dominance of humanities and languages in academic curricula in Germany has been considered responsible for the neglect of the potential embedded in international comparisons (Bos, Postlethwaite & Gebauer, 2010), but in terms of content this would only apply to mathematics and science, not to English as a foreign language and civic education. Others, finally (e.g., Klieme, 2011), have claimed that there was no political interest in recording and reacting to mediocre results of international comparisons at the time in question. This may, however, not be a very convincing explanation, given the immediate political reaction to Schultze's release of interstate comparisons without due consideration of sampling errors. In the summer of 1974, such figures were in the public, and the bitter controversies which ensued were indicative of strong differential interests rather than a general tendency to ignore them. Those federal states (Länder) credited with higher mean scores were unjustifiably attributing these numbers to superior educational policies, while those with lower scores were equally rash in complaining about shortcomings of the measurement. (See, for example, Walter Schultze's [1974b] attempt to clarify matters by



way of a short notice in a rather inconspicuous journal for a predominantly foreign [!] audience.)

In any event, the resultant political damage made German participation in IEA studies unviable for many years to follow (cf. Döbrich, 2002). In that sense, IEA-type research was, indeed, beginning to have political effects in the mid-1970s – at least in the sense that the respective findings were exploited as political ‘ammunition’, much beyond what would have been legitimate scientifically. It was to be expected, perhaps, that researchers, administrators and even policy makers were reluctant under these circumstances to risk their reputations in this minefield.

One of the two above-mentioned examples of researchers’ reference to the IEA approach is all the more noteworthy. Helmut Fend, then entrusted with a large-scale evaluation of the experimental Comprehensive Schools (*Gesamtschulen*), used some items from FIMS and the First International Science Study (FISS) as well as the entire English test from the IEA Study of English as a Foreign Language in his instrumentation of the assessment (Fend, 1982). The fact that he identified certain deficits in comprehensive schools particularly with regard to English as a foreign language did not generate a great deal of popularity for such tests among the adherents of this particular educational reform, at that time the mainstream among educators. At the same time, this data-based evaluation demonstrates that educational policy making, particularly if driven by reform motives, could not and cannot do without some form of empirical underpinning. Consider, for example, the host of *Gutachten*, i.e., supposedly evidence-based attempts to substantiate the recommendations of the German Educational Council (*Deutscher Bildungsrat*) which replaced the German Committee for Education in 1965. It is noteworthy here that the lack of trustworthy indicators of “what is possible” (Foshay et al., 1962) resulted in a set of recommendations considered illusionary by many and certainly unacceptable to contemporary ministers of finance (Hüfner et al., 1986).

The fear of repeating the bitter controversies of 1974 was henceforth equally present among researchers and policy makers. Some considered the likelihood of earning a scientific reputation by participating in IEA studies low, funding seemed unlikely, and even access to schools on the scale of anything approaching representativity could hardly be expected. Not until a decade later did a new generation of researchers engage in tentative efforts to reestablish contacts with IEA, albeit – for the time being – not at the level of nationwide assessments. Thus, a group from the Max Planck Institute for Psychology headed by Franz E. Weinert implemented and enhanced the IEA Classroom Environment Study (Anderson, Ryan & Shapiro, 1989) on the very modest scale of 39 classrooms from schools in and around Munich. A substantial number of important publications, in both German and international journals, emerged from this miniaturized study, and it definitely contributed to diverting attention from the fruitless debates over the proper structure of the German school system(s) to focusing on key elements of the quality of instruction. It is remarkable, however, that Weinert and his colleagues hardly ever, save for occasional comments and footnotes, referred to their project as part of the IEA Classroom Environment Study. To this day, it is much better known under the rather modest label “Munich Main School Study” (*Münchener Hauptschulstudie*: Helmke, Schneider & Weinert, 1986).

A second instance of joining a small regional study to a set of national samples within an IEA study was the “Hamburg Study of Achievement in Written Composition” (*Hamburger Aufsatzstudie*: Lehmann & Hartmann, 1987). It was restricted, as the name suggests, to the City of Hamburg, although both in terms of size and diversity the sample was not much inferior to some of the national samples. In this case, an impact on educational policy making could not really be expected. Nevertheless, the study served to facilitate technological transfer from an international project, operated at a state-of-the-art level, to the German context where in this domain only very limited evidence was available at that point (Lehmann, 1990). Related to this,

new approaches to assess student writing were – at least in principle – made available to the respective scientific and educational community in Germany. Although the International Study of Written Composition failed to demonstrate the comparability of essay ratings across countries and languages (see Purves, 1992), the Hamburg Study provided extremely valuable benchmarks when the collapse of the GDR facilitated this achievement survey in the former East Germany (Hartmann & Jonas, 1991).

### **The Stage of Unfolding Influence (1988 – 1999): Preparation of a Paradigm Shift in Research and Policy Making**

The IEA Reading Literacy Study (Elley, 1994) was the first in which full participation was achieved. In a first stage, all eleven West German *Länder* were convinced to participate-- whereupon the federal government agreed to fund the project. Then – with the sudden collapse of the East German government and the subsequent reunification – it became possible to extend the study to the schools in the five ‘new’ *Länder* and East Berlin which in 1990 had ‘joined’ the Federal Republic under Article 23 of the West German Constitution (*Grundgesetz*). The data collection took place in the spring of 1991, i.e., during the last school year of the former East German system of education. Given the unique opportunity of comparing two systems with different structures and governance, but significant cultural commonalities, it was decided to have two separate German samples, each satisfying the IEA requirements for country participation (Lehmann et al., 1995).

Consequently, the political significance of this study for Germany was highly concentrated on the intra-German comparisons. As a surprise to many, perhaps, it turned out that in both grades tested the two Germanys were remarkably similar in terms of mean achievement, variance and even the relationships between the individual social background and achievement (Lehmann et al. 1995). Even taking a ‘Composite Development Index’ into account, an adjustment that ben-

efited the former East, little was changed except teasing out a slight advantage for the East German students on the 'Documents' scale (Elley, 1992), explainable by the relatively higher emphasis on mathematics and science in the centralized East German curriculum. Although these results may have served to renounce false claims in later East vs. West debates over the relative merits of the respective systems of education, they could not influence the fundamental decision to reorganize education in the new *Länder* according to Western standards. This decision had already been taken, in 1990 (Article 37 of the Unification Treaty), linking all questions of detail to the accumulated resolutions of the Standing Conference of State Ministers of Education in the Federal Republic of Germany (KMK: *Ständige Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland*).

Opinions as to the political impact of the IEA Reading Literacy Study vary somewhat. Although it may be true that the findings, especially the disappointing achievement levels and the disclosed social disparities, went largely unnoticed outside a small circle of reading experts (van Ackeren & Klemm, 2000), other authors consider the Reading Literacy Study as having received broad public attention at least in the long run and thus representing, in conjunction with the Third International Mathematics and Science Study (TIMSS: Harmon et al., 1997; Baumert et al., 1997; Baumert, Bos & Lehmann, 2000), an important turning point in the relationship between IEA and policy makers in Germany (Jude & Klieme, 2010). In fact, the Reading Literacy Study had provided TIMSS with access to the field and, importantly, with sufficient federal funding to cover the middle school and the final school year populations.

It is true, however, that the main political effect is to be attributed to the TIMSS: although mediocre achievement levels and substantial social disparities in education had been a constant theme of IEA studies in Germany from the very beginning, this time public and political reactions to the publication of results during the winter of 1996/97

were aptly called a “shock”: “TIMSS-shock” – soon after replaced by the synonym “PISA-shock” – became a term which was then ubiquitous in public discourse, and certainly strong enough to force politicians to react.

KMK, the Standing Conference of Ministers of Education, was the primary agent from whom adequate policies were expected. It was, indeed, amazing to see how principles of the state rule, established half a century earlier and strongly defended by the Länder ever since, were abandoned within months in an effort to meet those expectations: The so-called Constance Resolution (*Konstanzer Beschluss*) of October 24, 1997, established that Germany should henceforth participate in internationally comparative studies with the aims of: securing high levels of quality in school-based education; guaranteeing equivalence of school-leaving certificates across states; and enhancing permeability across the extremely complex system of educational trajectories. ‘Monitoring’, ‘benchmarking’ and ‘evaluation’ became key terms of this new approach.

At about the same time, the methodological transfer based on IEA-type research began to make itself felt by facilitating state-level assessments. Modern techniques of measurement were implemented, enabling the investigators to study growth in complex school settings and to report findings back to schools in their endeavors to improve their own practices (see, e.g., Lehmann, Gänßfuß & Peek, 1999). Policy makers were encouraged by this kind of work to choose large-scale assessments as the starting point for improvements of instructions, school development and change at the system level. Hermann Lange, a renowned educational administrator and key person in the relevant KMK Committee on School Affairs, coined the term “empirical turn in educational policy-making” (*“empirische Wende der Bildungspolitik”*: Lange, 1999).

This phrase aptly marks both the theoretical and the political paradigm shift initiated by IEA studies in Germany. From then on, educa-

tional policy making without reference to internationally established norms was no longer a serious option.

### **Consolidating the Influence of IEA under the New Paradigm (from 2000 Onwards)**

At the political level, there were even farther-reaching repercussions of TIMSS. In the year 2000, a broad alliance involving the federal and state governments, important societal groups such as trade unions and employers' associations, educators, churches, apprentices and university students, set up the Educational Forum (*Forum Bildung*). It acted under the patronage of the President of the Federal Republic and passed a number of important resolutions intended to guide improvements in educational policy (Arbeitsstab Forum Bildung, 2001).

One of the most consequential outcomes was the decision to define standards for various domains and school leaving certificates, having the option in mind to enlarge that program by also covering important transition points in the educational system, such as the point(s) of allocating students to the various forms of secondary schooling. Since the objective of working with evidence-based definitions implied massive efforts in terms of data collection and analysis, the KMK set up and endowed (jointly with Humboldt University, Berlin) the Institute for Quality Management in Education (IQB: *Institut zur Qualitätsentwicklung im Bildungswesen*). Meanwhile, many of the Länder have founded their own agencies for quality management, charged with conducting large-scale assessments and related activities, often in the form of a census of achievement at particular grade levels in the entire state. To a high degree, this is a reflection of the fact that ever since the dissemination of TIMSS results in 1997 and a few impulsive reactions, there was an almost unconditional support for large-scale assessments in the country. Since, however, the data collection and analysis, let alone the interpretation and feedback were not always performed with the required prudence by the state-sponsored

institutions, some concerned actors are beginning to have reservations with regard to an unlimited confidence in such activities. Clearly, efforts are required to foster the quality of these activities as well, not least by training highly qualified personnel. An important step in such capacity building was facilitated by the decision of the German Research Association (DFG: *Deutsche Forschungsgemeinschaft*) to fund the international Coordination of the IEA Civic Education Study of 1999/2000 (Torney-Purta et al., 2001; Amadeo et al., 2002). To foster technology transfer was a primary motive for that decision, which was fully justified by the highly efficient and innovative involvement of the IEA Data Processing Center (DPC) in the international data processing and analysis for this study (Schulz & Sibberns, 2004).

Universities and the big research institutes belonging to the Leibniz Society have responded to these needs by setting up departments of (empirical) educational research and by creating graduate schools, doctoral programs and research clusters. Workshops introducing novices to state-of-the art computer technology have become quite common, inviting research groups whose work so far has been, perhaps, characterized by more qualitative techniques to adopt the new paradigm. Characteristically, this is the case in areas where evidence-based applied pedagogical content knowledge (didactics) is to be developed. Here, too, the DPC in Hamburg has earned a unique reputation by contributing to these activities, not least by joining forces with the renowned Educational Testing Service of Princeton (NJ) in founding the IEA-ETS Research Institute (IERI).

It may be mentioned here that the IEA Data Processing Center (DPC) in Hamburg has become the prime contractor for undertaking virtually all government-authorized large-scale assessments in Germany, securing the required quality in the production and sometimes also in the analysis of the respective data sets. The impartial provision of evidence is, indeed, a necessary condition for rational policy making. Those instances where such impartiality has been jeopardized by contracting commercial firms that were sometimes even inexperienced in

the field of education clearly demonstrate the necessity to adhere strictly to a new paradigm. They also show, however, that educational policy makers feel obliged to convey the impression of acting on the basis of empirical evidence, however inadequate it may be.

It is an indication of the persisting impact of IEA on German educational policy-making that the KMK has committed the Länder to regular participation in the IEA studies geared towards primary education. The Institute for School Development (IFS: *Institut für Schulentwicklungsforschung*) at the Technical University of Dortmund has participated in PIRLS 2001 and continues to do so (Bos et al., 2003, 2007), an interesting side observation being that once again the reference to the large international IEA Study is somewhat obscured by choosing and using the German acronym IGLU (*Internationale Grundschul-Lese-Untersuchung*). It is true, however that due recognition is given in prominently placed footnotes, and certainly, little public attention and political effect would be achieved without the essential benchmarks derived from the international comparisons.

It was a brilliant maneuver to include the TIMSS primary component in the IGLU 2000, thus establishing a baseline for the respective cycle of TIMSS studies to come. Unfortunately, analogous attempts to embark on the TIMSS-Advanced strand have failed, allegedly – but not without some plausibility – for a lack of public funds. Despite such occasional setbacks, it seems fair enough to conclude that educational policy makers in Germany have once and for all acknowledged the indispensable value of international benchmarks for educational decisions.

The present description of IEA influence on educational policy making in Germany has not been complete in the sense of discussing every instance of German involvement in IEA projects. The – once again, ‘half-hearted and intermittent’ – participation in IT-related studies could have been analyzed, just as the many unsuccessful efforts to acquire funds for a German component to the International Civic and



Citizenship Education Study (ICCS 2009). Yet, in contrast to these examples, the German contribution to the IEA Teacher Education and Development Study in Mathematics (TEDS-M: Tatto et al., 2008) has initiated a whole line of national project continuations which reflect the political awareness of another prerequisite of improvements in learning, namely the professional preparation of teachers. As teacher education is currently (2011) in a state of rapid change triggered by the so-called Bologna Process, the expected findings will be timely, indeed, for those who are responsible for respective decisions.

## References

- Amadeo, J.-A., Torney-Purta, J., Lehmann, R., Husfeldt, V. & Nikolova, R. (2002). *Civic Knowledge and Engagement. An IEA Study of Upper Secondary Students in Sixteen Countries*. Amsterdam: IEA.
- Anderson, L. W., Ryan, D. W., & Shapiro, B. J. (1989). *The IEA Classroom Environment Study*. Oxford: Pergamon Press.
- Arbeitsstab Forum Bildung (Ed.). (2001). *Empfehlungen des Forum Bildung*. [Recommendations of Forum Education]. Bonn: Bund-Länder-Kommission für Bildungsplanung und Forschungsförderung.
- Baumert, J., Lehmann, R., Lehrke, M. Schmitz, B., Clausen, M., Hosenfeld, I., Köller, O., & Neubrand, J. (1997). *TIMSS - Mathematisch-naturwissenschaftlicher Unterricht im Internationalen Vergleich*. [TIMSS - International Comparison of Mathematics and Science Instruction]. Deskriptive Befunde. Opladen: Leske + Budrich.
- Baumert, J., Bos, W., & Lehmann, R. (Eds.). (2000). *TIMSS/III. Dritte Internationale Mathematik- und Naturwissenschaftsstudie - Mathematische und naturwissenschaftliche Bildung am Ende der Schullaufbahn*. [TIMSS/III - Third International Mathematics and Science Study - Mathematics and Science Achievement at the End of Secondary Schooling]. 2 vols. Opladen: Leske + Budrich.
- Bos, W., Lankes, E.-M., Prenzel, M., Schwippert, K., Valtin, R. & Walther, G. (Eds.). (2003). *Erste Ergebnisse aus IGLU. Schülerleistungen am Ende*

- der vierten Jahrgangsstufe im internationalen Vergleich.* [First Results of the IGLU Study (*i.e.*, the German contribution to the IEA PIRLS). International Comparison of Student Achievement at the End of Grade 4]. Münster/New York: Waxmann.
- Bos, W., Hornberg S., Arnold, K.-H., Faust, G., Fried, L., Lankes, E.M., Schwippert, K., & Valtin, R. (Eds.). (2007). *IGLU 2006: Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich.* [IGLU 2006 (*i.e.*, the German contribution to the IEA PIRLS 2006). International Comparison of Primary Students' Reading Comprehension]. Münster/New York: Waxmann.
- Bos, W., Postlethwaite, T. N., & Gebauer, M. M. (2010). Möglichkeiten, Grenzen und Perspektiven internationaler Schulleistungsforschung. [Opportunities, Limitations, and Perspectives of International Research on Educational Achievement]. In R. Tippelt & B. Schmidt (Eds.): *Handbuch der Bildungsforschung.* [Handbook of Educational Research]. 3., durchgesehene Auflage. [3rd Rev. Ed.]. Wiesbaden: Verlag für Sozialwissenschaften, 275-295.
- Carroll, J. B. (1975). *The Teaching of French as a Foreign Language in Eight Countries.* Stockholm/New York: Almqvist & Wiksell/John Wiley & Sons (International Studies in Evaluation V).
- Comber, L. C., & Keeves, J. P. (1973). *Science Education in Nineteen Countries.* An Empirical Study. Stockholm/New York: Almqvist & Wiksell/John Wiley & Sons (International Studies in Evaluation I).
- Dahrendorf, R. (1965). *Bildung ist Bürgerrecht. Plädoyer für eine aktive Bildungspolitik.* [Education is a Civil Right. A Treatise in Favor of an Active Educational Policy]. Hamburg: Nannen-Verlag. [Key passages translated into English in S. Melnik, & S. Tamm (Eds.). (2008). *Liberal Readings on Education.* Berlin: liberal Verlag].
- Döbrich, P. (Ed.). (2002). *Qualitätsentwicklung im naturwissenschaftlichen Unterricht.* Fachtagung am 15. Dezember 1999. [Quality Development in Mathematics and Science Instruction. Workshop on December 15, 1999]. Frankfurt/Main: Gesellschaft zur Förderung Pädagogischer Forschung (Materialien zur Bildungsforschung, Bd. 7).

- Elley, W. B. (1992). *How in the world do students read? IEA Study of Reading Literacy*. Amsterdam: IEA.
- Elley, W. B. (Ed.). (1994). *The IEA Study of Reading Literacy: Achievement and Instruction in Thirty-Two School Systems*. Oxford: Pergamon Press (International Studies in Educational Achievement, vol. 11).
- Fend, H. (1982). *Gesamtschule im Vergleich*. Bilanz der Ergebnisse des Gesamtschulversuchs. [Evaluating Comprehensive Schooling. Assessing the Results of the Experimental Introduction of Comprehensive Schools]. Weinheim/Basel: Beltz.
- Foshay, A.W., Thorndike, R.L., Hotyat, F., Pidgeon, D.A., & Walker, D.A. (1962). *Educational Achievement of Thirteen-Year-Olds in Twelve Countries*. Hamburg: UNESCO Institute for Education.
- Harmon, M., Smyth, T. A., Martin, M. O., Kelly, D. L., Beaton, A. E., Mullis, I. V. S., Gonzales, E. J., & Orpwood, G. (1997). *Performance Assessment in IEA's Third International Mathematics and Science Study*. Chestnut Hill, MA: Boston College.
- Hartmann, W., & Jonas, H. (1991). *Deutschunterricht im Umbruch. Die Aufsatzstudie-Ost von 1991*. [Transforming the Teaching of German. The East German Study of Written Composition of 1991]. Frankfurt/Main: Peter Lang.
- Helmke, A., Schneider, W., & Weinert, F. E. (1986). Quality of instruction and classroom learning outcomes: The German contribution to the IEA Classroom Environment Study. *Teaching and Teacher Education*, 2, 1-18.
- Hüfner, K., Naumann, J., Köhler, H., & Pfeffer, G. (1986). *Hochkonjunktur und Flaute. Bildungspolitik in der Bundesrepublik Deutschland 1967-1980*. [Boom and Depression. Educational Policy-Making in the Federal Republic of Germany 1967 - 1980]. Stuttgart: Klett-Cotta.
- Husén, T. (Ed.). (1967). *International Study of Achievement in Mathematics. A Comparison of Twelve Countries*. 2 vols. Stockholm/New York: Almqvist & Wiksell/John Wiley & Sons.
- Jude, N., & Klieme, E. (2010). Das Programme for International Student Assessment (PISA). In E. Klieme, C. Artelt, J. Hartig, N. Jude, O. Köller,

- M. Prenzel, W. Schneider, & P. Stanat (Eds.). *PISA 2009. Bilanz nach einem Jahrzehnt*. [PISA 2009. Taking Stock after a Decade]. Münster/New York: Waxmann, 11-21.
- Klieme, E. (2011). Bildung unter undemokratischem Druck? Anmerkungen zur Kritik der PISA-Studie. [Education under Undemocratic Pressure? Some remarks on the Criticism of the PISA Study]. . In St. Aufenanger, F. Hamburger, L. Ludwig & R. Tippelt (Eds.) *Bildung in der Demokratie*. Beiträge zum 22. Kongress der DGfE. [Education and Democracy. Contributions to the 22nd Convention of the German Educational Research Association]. Band 2. Opladen/Farmington Hills: Barbara Budrich (in print).
- Lange, H. (1999). Qualitätssicherung in Schulen. [Securing Quality Schools]. *Die Deutsche Schule*, 91, (2), 144-159.
- Lehmann, R. H. (1990). Aufsatzbeurteilung - Forschungsstand und empirische Daten. [Grading Essays - State of Research and Empirical Data]. In K. Ingenkamp, & R.S. Jäger (Eds.), *Tests und Trends*. 8. Jahrbuch der Pädagogischen Diagnostik. [Tests and Trends. 8th Yearbook of Educational Measurement]. Weinheim/Basel: Beltz, 64-94.
- Lehmann, R. H., & Hartmann, W. (1987). *The Hamburg Study of Achievement in Written Composition*. National Report for the IEA International Study of Achievement in Written Composition. 2 vols. Hamburg (School of Education).
- Lehmann, R. H., Peek, R., Pieper, I., & von Stritzky, R. (1995). *Leseverständnis und Lesegewohnheiten deutscher Schüler und Schülerinnen*. [Reading Comprehension and Reading Habits of German Students]. Weinheim /Basel: Beltz.
- Lehmann, R. H., Gänsfuß, R., & Peek, R.(1999). *Aspekte der Lernausgangslage und der Lernentwicklung von Schülerinnen und Schülern an Hamburger Schulen - Klassenstufe 7*. Bericht über die Untersuchung im September 1998. [Aspects of Initial Achievement Levels and Learning of Students in Schools in the City of Hamburg - Grade 7. Report on the 1998 Census]. Hamburg: Behörde für Schule,

- Jugend und Berufsausbildung. [<http://www.hamburger-bildungsserver.de/welcome.phtml?unten=/schulentwicklung/lau/lau7/>]
- Lewis, E. G., & Massad, C. E. (1975). *The Teaching of English as a Foreign Language in Ten Countries*. Stockholm/New York: Almqvist & Wiksell/John Wiley & Sons (International Studies in Evaluation IV).
- Oppenheim, A. N. (1976). Neue Grundlagen für die Demokratie in Deutschland: Fiktion oder Wirklichkeit? [New Foundations for Democracy in Germany: Fiction or Reality?]. *Neue Sammlung*, 16, 47-66.
- Picht, G. (1964). *Die deutsche Bildungskatastrophe*. Analyse und Dokumentation. Mit einem Vorwort. [The German Educational Disaster. Analysis and Documentation, with a Preface]. Olten/Freiburg im Breisgau: Walter-Verlag.
- Postlethwaite, T. N. (1995). International empirical research in comparative education. An example of the studies of the International Association for the Evaluation of Educational Achievement (IEA). *Tertium comparationis*, 1, 1-19.
- Purves, A. C. (1973). *Literature Education in Ten Countries*. An Empirical Study. Stockholm/New York: Almqvist & Wiksell/John Wiley & Sons (International Studies in Evaluation II).
- Purves, A. C. (1992). A Comparative Perspective on the Performance of Students in Written Composition. In A.C. Purves (Ed.), *The IEA Study of Written Composition II: Education and Performance in Fourteen Countries*. Oxford: Pergamon Press (International Studies in Educational Achievement, vol. 6), 129-152.
- Schultze, W., & Riemenschneider, L. (1967). Eine vergleichende Studie über die Ergebnisse des Mathematikunterrichts in zwölf Ländern. [A comparative study of the success of mathematics instruction in twelve countries]. Frankfurt am Main: *Mitteilungen und Nachrichten des DIPF*, No. 46/47, 1-34.
- Schultze, W. (1974a). Die Leistungen im naturwissenschaftlichen Unterricht in der Bundesrepublik im internationalen Vergleich. [Science Achievement in the Federal Republic in an International

- Comparison]. Frankfurt/Main: *Mitteilungen und Nachrichten des DIPF* (Sonderheft): DIPF.
- Schultze, W. (1974b). In Naturwissenschaften kein Musterschüler. Die Leistungen im naturwissenschaftlichen Unterricht in der Bundesrepublik im internationalen Vergleich. [Not a top student in science. International comparison of science achievement in the Federal Republic]. In: *Bildung und Wissenschaft*, No. 13, 185-187. Bad Godesberg: Inter Nationes e. V.
- Schultze, W. (1975). Die Leistungen im Englischunterricht in der Bundesrepublik im internationalen Vergleich. [International comparison of achievement in English as a foreign language in the Federal Republic]. Frankfurt/Main: *Mitteilungen und Nachrichten des DIPF* (Sonderheft): DIPF.
- Schulz, W., & Sibberns, H. (Eds.). (2004). *IEA Civic Education Technical Report*. Amsterdam: IEA.
- Seibel, H. D., & Wolf-Seibel, H. R. (1976). Vergleichende Bildungsforschung. Replik auf Hans Freudenthals Diskussion der IEA-Untersuchung. [Comparative educational Research. A reply to Hans Freudenthal's discussion of the IEA studies]. *Zeitschrift für Pädagogik*, 22, 103-107.
- Tatto, M. T., Schulle, J., Senk, S., Ingvarson, L., Peck, R., & Rowley, G. (2008). *Teacher Education and Development Study in Mathematics (TEDS-M): Conceptual framework*. East Lansing, MI: Teacher Education and Development International Study Center, College of Education, Michigan State University.
- Tenorth, H.-E. (2010). Empirische Schulforschung aus bildungshistorischer Sicht. [Empirical research on teaching from the perspective of Educational History] .In J.D. Gauger, & J. Kraus (Eds.), *Empirische Bildungsforschung - Notwendigkeit und Risiko*. Sankt Augustin/Berli: Konrad-Adenauer-Stiftung, 7-20.
- Thorndike, R. L. (1973). *Reading Comprehension Education in Fifteen Countries*. An Empirical Study. Stockholm/New York: Almqvist &

- Wiksell/John Wiley & Sons (International Studies in Evaluation III).
- Torney, J. P., Oppenheim, A. N. & Farnen, R. F. (1975). *Civic Education in Ten Countries. An Empirical Study*. Stockholm/New York: Almqvist & Wiksell/John Wiley & Sons (International Studies in Evaluation VI).
- Torney-Purta, J., Lehmann, R., Oswald, H. & Schulz, W. (2001). *Citizenship and Education in Twenty-eight Countries. Civic Knowledge and Engagement at Age Fourteen*. Amsterdam: IEA.
- van Ackeren, I. (2002). Von FIMS und FISS bis TIMSS und PISA. Schulleistungen in Deutschland im historischen und internationalen Vergleich. *Die Deutsche Schule*, 94 (2), 157-175.
- van Ackeren, I. & Klemm, K. (2000). TIMSS, PISA, LAU, MARKUS und so weiter - Ein aktueller Überblick über Typen und Varianten von Schulleistungsstudien. *Pädagogik*, 52, (12), 10-15.

## CHAPTER 18

# How IEA Influenced the Education System in Hungary

*Sándor Brassói & Judit Kádár-Fülöp*

### How Hungary Joined the IEA

Hungary joined the IEA in 1968. It was a year of political challenges on both sides of the Iron Curtain: the year of the first big oil crisis, which changed the economic outlooks of eastern European countries, the year of student riots in western Europe, the year of the Prague Spring, and the tenth anniversary of the execution of Imre Nagy, prime minister during the Hungarian revolution of 1956. The Vietnam War was raging, but the Nuclear Non-Proliferation Treaty was ready for signing. In the race for information technology development, the U.S. surpassed the Soviet Union and had a much greater human potential for further development than her competitor. How far global and European politics influenced the lenience of communist politicians (in Budapest and in Moscow), who allowed Hungary to start building research contacts with western countries, is not known. It seemed however, that Hungary was considered an experimental field in this respect. The Kádár version of socialism--nicknamed "gulasch socialism"--seemed a possible solution: some lenience in small things in exchange for full obedience in the big things served Moscow's political ambitions without further destroying its reputation. The government attempted to reconcile political demands from Moscow with researchers' needs for more cooperation with their colleagues in the West, so that although the Hungarian Ministry of Foreign Affairs continued to control western contacts at both individual and institutional level, the political grip on social scientists was - albeit hardly perceptibly - loosened.

On the other side of the Iron Curtain, those who were in charge of the IEA appeared interested in involving at least some eastern European



countries in the comparative assessment of student learning outcomes. After the exodus in 1956, there was a meagre but constant emigration of academics from eastern Europe whose competence suggested that education behind the Iron Curtain was worth studying. Furthermore, the IEA pilot study (Foshay et al., 1962) revealed that gender differences in mathematics and science show a different pattern in communist Yugoslavia than in western countries, i.e. girls perform better or at least not worse than boys. The Six Subject Survey appeared a good occasion to enlarge the IEA community to include eastern European institutions. Around that same time, some Hungarian educational researchers were already advocating national assessment as an alternative to using teacher awarded grades in the analysis of educational achievement. Árpád Kiss, Head of the Department of Didactics in the National Institute for Public Education, launched the first large-scale assessment in 1967 – such a thing had been considered a sacrilege just a few years before – and invited Zoltán Báthory to manage the study. So, when Neville Postlethwaite, the managing director of IEA, arrived in Hungary in the same year, he met researchers both in Budapest and in Szeged who were interested in an empirical comparative study and who were ready to learn large-scale assessment methodology. They also had some experience in the technical difficulties and the amount of work involved. So the technological and professional advantages of joining the IEA were well understood: the collaboration with top experts in educational measurement, and – as in those years social science researchers in Hungary rarely had any computers at all – the level of technology available, guaranteed professional quality and feasibility.

The report of the first mathematics study (Husén, 1968) came out in 1968 and it settled most of the doubts that some Hungarian experts had expressed on the validity and reliability of international comparisons. The main concern regarding the curriculum differences and their impact on student achievement was to be treated by the Opportunity To Learn questionnaire in the 1970 study. A second concern of Hungarian experts was the use of multiple choice tests, with

which Hungarian students had no experience. However, it was hoped that the worldwide acceptance of the advanced sampling and measurement techniques used in the IEA mathematics study would remove scruples about comparability in Hungary. The first interviews with Neville Postlethwaite, who had a special talent for explaining complicated technical ideas with great clarity and in a simple way, convinced Árpád Kiss that this was a unique opportunity to make a great step forward in empirical research in education. His colleague, Zoltán Báthory, had proved himself an able coordinator in the first large-scale national assessment which was underway at the time, and he was also very interested in cooperation with leading researchers in empirical research. Judit Kádár-Fülöp, who participated in the talks as interpreter to Neville Postlethwaite, could see an opportunity to become a member of the study team. All were eager to join the innovative and challenging Six Subject Survey.

Negotiations with the Ministry of Education and the Ministry of Foreign Affairs started at the end of 1967. Árpád Kiss, who was said to be in possession of an IEA letter with a letterhead on which the Academy of Pedagogy in Moscow was listed, succeeded with this letter in hand to persuade the competent party and ministry officials that there was no political risk for Hungary in joining the IEA. Such a letterhead may have existed for a single conference or for a very short period, because neither the Soviet Union nor any of the east European countries other than Hungary had joined IEA before the collapse of the eastern European communist regimes. Although there were attempts made by Polish and Rumanian institutions to carry out some of the IEA surveys, neither country produced a completed and published dataset.

### **IEA Studies in Hungary and their Impact on Education before the Millennium**

*This section summarizes the effects of IEA participation in educational research and development during the decades when IEA was the only agent*

*in the field of international student assessment. It shows how the Six Subject Survey fit the policy environment of the 70s and how it inspired reforms in curriculum development, textbook development and the practice of monitoring student achievement in Hungary.*

The first IEA study in which Hungary participated was the Six Subject Survey of 1970-71 that examined reading comprehension and science at the primary, lower secondary and upper secondary level. It also tested English as a Foreign Language at the upper secondary level. Torsten Husén, the Chair of IEA at the time, and Neville Postlethwaite, who coordinated the Six Subject Survey, helped the Hungarian team in every possible way: with information as well as by offering financial support to attend international meetings (there was no way to get money for travelling from Hungarian resources at the time).

The Hungarian results were good in science and mathematics, but rather poor in reading--particularly at the primary and the upper secondary level. They were very poor in English as a foreign language. The Hungarian media were silent about the study and its results - as was usually the case with educational or technological news that would not show the superiority of the communist system. However, the publications in professional periodicals (Báthory et al., 1976; Kiss et al., 1979) were widely read and raised awareness of the strong and weak sides of the Hungarian education system.

After the first sceptic reactions from politicians and educationists who did not believe that knowledge and skills could be tested in an internationally comparable way (particularly not with multiple choice tests), the Six Subject Survey results opened a floodgate of debate on survey research and its possible role in influencing educational policy. Thanks to the mixed results and the good news about science achievement, political repercussions were subdued, which was a sign of silent consent for the National Institute of Public Education to continue collaboration with IEA.

The interest in comparative studies of students' achievement and their

school conditions was well founded. The Coleman Report (1966), which provided statistical evidence for the strong influence of home background on school achievement, raised awareness of the difficulties of educational policy to enhance social mobility. Equity in access to education was an acute political problem in Hungary as well. Egalitarianism and positive discrimination of the youth from the 'working classes' in accessing higher education, did not automatically lead to the desired educational mobility. This was politically alarming and threatened the smooth takeover of power by the new "working class aristocracy." Therefore, the educational factors of social mobility and the sociocultural factors of educational success came to the focus of research interest. The IEA study showed that the Hungarian educational system, with its egalitarian philosophy, was actually less successful in promoting educational mobility than the Swedish or the Finnish educational system (Báthory, 1992). The relatively high percentage of variance attributable to differences among schools and the strong correlation between the average SES level of the school and school performance in comparison to Sweden surprised education policy makers who had believed that administrative control over school admission, uniform curriculum and textbooks were sufficient to close the cultural gap between middle class and working class children. However, an important group of Hungarian sociologists, who were studying the school's impact on educational mobility, found their own survey findings corroborated by the Hungarian IEA results. They welcomed the empirical results, which supported their suggestions to provide better and more substantive support to children from culturally deprived families.

Thanks to Zoltán Báthory, who became Head of the Evaluation Centre of the National Institute for Public Education in the late 1970s, Hungary remained a regular participant in the IEA studies after the Six Subject Survey as well (cf. Table 1). At present, Hungary participates in the TIMSS and PIRLS surveys.

In spite of our zealous participation in several IEA surveys, their

**Table 1.** Hungary's Participation in IEA projects

Name of the Project	Year of data collection	Name of the National Research Coordinator (NRC)	Member of the IEA Governing Board
IEA Six Subject Study	1970-71	Zoltán Báthory, Judit Kádár-Fülöp	Zoltán Báthory
SIMS	1980-81	Gusztáv Hábermann	
Classroom Environment Study	1981-1983	András Joó	
SISS	1983-84	Zoltán Báthory, Péter Vári	
International Study on Composition Skills	1985	Judit Kádár-Fülöp	
COMPED	1989-1992	Péter Vári	
Second IEA Study on Reading Literacy	1991	András Láncki, Judit Kádár-Fülöp, Péter Vári	
TIMSS	1995	Péter Vári	Péter Vári
SITES	1999	Péter Vári	
TIMSS first repetition (TIMSS-R)	1999	Péter Vári	
CIVED Modul 1	1999	Zsuzsa Mátrai	
PIRLS 2001	2001	Péter Vári	Sándor Brassói
TIMSS 2003	2003	Péter Vári	
PIRLS 2006	2006	Ildikó Balázsi	
TIMSS 2007	2007	Ildikó Szepesi	

results had little direct impact on education policy or instruction in Hungary. The educational focus remained on input until 2000: curriculum and textbook reforms dominated the 1980s and the 1990s. It must be mentioned, however, that IEA also promoted international learning in the field of curriculum development. By organizing international seminars for national teams of policy makers, evaluation experts and curriculum experts, the IEA research group succeeded in building an international “club of policy makers” who understood the

advantages of evidence-based policy. The Hungarian experts (Zoltán Báthory, Endre Ballér, Péter Szebenyi, Tamás Varga) who participated in the Gränna seminar in 1971 (see the chapters by John Keeves and by Kimmo Leimu) became leading figures in Hungarian educational research and policy in the subsequent decades. What they learned in Gränna shaped the curriculum reforms of the 1980s and 1990s in Hungary.

IEA also had an impact on innovation and empirical research in education. It is worth mentioning, for example, that the poor results in the 1970 reading survey, which were first received with scepticism by primary teachers unfamiliar with objective testing methods, generated a fruitful discussion on how comprehension skills should be developed and on the role of silent reading. At the same time, the poor Hungarian results provided a welcome argument for the advocates of textbook choice – until then there had been only one series of textbooks for all basic schools. Several proposals to develop new textbooks and alternative teaching methods were given a green light after the IEA experience. This was the advent of a textbook market: in the 1980s, several primers and textbooks for the elementary grades were developed by different groups of textbook writers. By using them, teachers became involved in learning different teaching strategies. The results of these efforts bore fruit, as was shown later in the PIRLS studies.

It should also be mentioned that objective testing started to gain ground in Hungarian schools. Maybe a little bit too quickly and mindlessly in some areas, as we thought at that time. However, in science teaching, where test development was undertaken by professional test designers, objective testing became a very useful tool in summative evaluation.

A major impact of IEA was the establishment of a system of national assessments inspired by IEA studies. After the first pilot national reading survey (Kádár-Fülöp, 1985), a regular monitoring of educational achievement was established in 1986 for reading, mathematics, science

and information technology. The national assessment surveys between 1986 and 1999 were connected to the IEA surveys and used the same samples; they followed up on trends of student achievement until 1999 (Vári, 1997, 1999).

In the 1990s, the driving force behind IEA participation and the national assessment studies was Péter Vári. Together with Zoltán Báthory, who was Deputy State Secretary for Public Education in the Ministry of Education between 1994 and 1998, Vári set up a system of financial support connected to both the IEA participation and the monitoring activity of national assessments.

Since 1996, the survey results have been fed into the Report on Public Education, a publication reporting on the state of education and its trends every three years. The regular use of survey data in reporting on the state of education confirmed the need for national and international surveys. At the end of the 1990s, the government established a regular budget line for international and national assessments. This was a major advancement--until then there had been no stable financial commitment by the government to carry out large-scale international or national educational assessments (including participation in IEA projects), meaning that the Evaluation Centre had to negotiate for each survey individually.

### **The impact of large-scale surveys on Hungarian education policy in the new millennium**

*This section discusses the major impacts of the international student achievement studies on Hungarian education policy after 2000. First the establishment of the National Assessment of Basic Competences will be summarized, followed by discussion of two topics, dissemination of achievement outcomes and capacity building. Finally, the impact on raising assessment standards will be mentioned.*

The IEA studies were the first to break down some of the barriers

between political ideologists and realists in educational policy. Although it has been a long struggle with many victories and defeats, today we think that evidence-based policy has a good chance to win. The public view is that education should be outcome-oriented rather than controlled by curricula and ideologies, and the international studies have speeded up this change in thinking.

The first results of OECD's Programme for International Student Assessment (PISA 2000) increased public awareness of international survey results and in some countries caused what was termed a "PISA shock". As a consequence, governments were pressed to design policy tools that would help improve results, reduce the number of students with very poor competences, and reduce the achievement gap between schools and between students. In Hungary, too, some of the policies developed in the past decades are strongly related to international assessment practices. Of these, four will be mentioned: i) the establishment of the National Assessment of Basic Competences, ii) dissemination policy regarding survey results and knowledge of how to use them, iii) development of an assessment infrastructure (capacity building), and iv) establishment of methodological standards for large-scale assessments.

### *The National Assessment of Basic Competences*

Recognizing that national reporting of sample survey data cannot by itself make a change, in 2000 the Hungarian Ministry of Education decided to set up a national assessment system that enables schools and stakeholders – including the local municipalities, the parents and the school staffs – to follow up on student progress within their area of competence.

After a pilot study in the school year 2001/2002, the National Assessment of Basic Competences replaced the national assessment programme beginning in the school year 2002/2003. Unlike the former national assessment surveys, which were sample-based surveys, the National Assessment of Basic Competences was designed as a full-



scale annual assessment of all students in Grades 6, 8 and 10, testing reading literacy and mathematical literacy towards the end of the school year. The assessment is based on a testing framework similar to those used in the PISA and TIMSS studies. The tests are administered by independent test administrators. All Grade 8 and Grade 10 tests and 50% of the Grade 6 tests are marked by the National Assessment Centre. The schools receive a scoring guide to mark the tests remaining in the school (the other 50 %). All schools have access to their own database so that they can further analyze their data, and there is also a downloadable software available that enables them to do this easily. In addition, all schools receive a school report that allows them to compare their results with the national and the county level means, as well as with schools whose student population has a similar sociocultural background. Seminars are organized to train school personnel and interested experts in using the software for data analysis. Since 2010, schools have had access to their students' previous survey results administered in another school, although, the student's individual data cannot be used for selection purposes.

There is an increasing use of school data by the schools themselves and by other stakeholders including researchers and parents. Results are discussed in more and more schools by the teaching staff and there are increasingly more measures taken to improve the reading and mathematical competence of students. Whereas in the beginning schools 'allowed' the survey but did not pay much attention to it, now they keenly follow change from one year to the next. Most schools have recognized that in the absence of a national school inspectorate, this is the only possible way for them to get objective feedback on how well they fulfil their objectives and how their students' level of achievement compares with that of students in other schools.

The availability of high quality data on learning outcomes of 14-year-olds enabled researchers to use these as a benchmark in studying the educational career of the 10,000 students sampled for the first Hungarian Life Course Study launched in 2006. The longitudinal

study marks its fifth year in 2011 and has yielded a very rich source of information on the social and educational state and perspectives of our teenagers.

### *Dissemination and Further Use of Survey Data*

Until 2001 the publication of the results of IEA and national assessment surveys was limited to papers in professional periodicals and eventually monographs depending on the author's research interest. The tri-annual *Report on Public Education*, first published in 1997, was the first general education policy publication reporting on international and national assessment outcomes in both an international and a national context. However, none of the databases emerging from the surveys were developed into a data product or were available to researchers. Furthermore, there was practically no media publicity for any of the surveys.

Since 2001, however, there has been a significant breakthrough in this field as well. The appearance of PISA on the student survey palette, drew significant media attention to educational outcomes worldwide. PISA raised public awareness that educational success is related to educational policy and the public debate it generated through the media made it clear to policy makers that PISA *put education policies to the test*.

Also at this time, the Ministry of Education began to be interested in organizing a press conference on the occasion of the release of the IEA and PISA surveys. Right before the international press conference, where the outcomes of an international achievement study are presented, the national media are fully informed on the specific objectives and significance of a given study.

Recognizing the value of public support for policy development, the National Assessment Centre responsible for international surveys was required to prepare a detailed executive summary for the press events. Over the years these executive summaries have become more and

more informative and rich in content, containing tables, figures and analyses that are interesting to the Hungarian public. It has become common practice to provide journalists with the executive summary of both the national and the international assessments.

Accordingly, media coverage has grown very considerably since 2001, creating public debate and an opportunity for leading experts in education and educational policy to discuss policy alternatives and educational values in the electronic media.

It has become standard procedure to set up a website for each survey that includes the assessment frameworks, sample items and the related reports, as well as links to the international homepage <http://oecd-pisa.hu>, <http://www.pirls.hu>. <http://www.timss.hu>. This ensures transparency and public access to all information related to the surveys.

Further to this, the National Assessment Centre compiles publications on the international survey results in Hungarian and English; these are electronically available for the general public and are available in print for the participating schools. The Centre also organizes in-service training courses and conferences for headmasters, representatives of school maintainers, and researchers. <http://www.kompetenci-amer.hu>

### *Development of an Institutional Framework for Assessment*

The research group implementing the IEA studies in Hungary and organizing national assessments in the 1970s and 1980s grew out of the Department for Didactics of the National Institute for Public Education led by Árpád Kiss and later by Zoltán Báthory. Subsequent to 1989-- in the 1990s-- the National Institute for Public Education was reorganized several times. Some research and development functions were commissioned to another institution, and there were also mergers of some institutes, resulting in the establishment of the National Assessment Centre. These developments brought beneficial changes for survey research and for the National Assessment Centre itself.

Péter Vári, taking on Zoltán Báthory's role as head of the National Assessment Centre in 1994, recruited and trained a permanent survey team of young experts who were able not only to collect but also to analyze and report data. From this time onwards, the Centre has had a specialist team whose main function is to develop, implement and report on large-scale assessments in education. The survey team survived the institutional reorganizations without losing its professional identity as a community or its personnel.

In 2003, the National Assessment Centre was incorporated into the National Office of Education attached to the Ministry of Education. The Office takes a key role in accreditation and quality monitoring in the public education system.

Within the Office, the Centre has remained a separate department headed by Ildikó Balázs. The Centre implements the TIMSS and PIRLS surveys and the PISA surveys. Beyond this, the Centre develops and implements the annual National Assessment of Basic Competences (NatABC) in Grades 6, 8, and 10 and reports on both of the international surveys and on the NatABC. The Centre also develops software for schools for school-level data analysis.

The National Assessment Centre also has a wide range of publication activities: besides printed reports, it publishes assessment frameworks, sample items and specific reports on its websites ([www.oecd-pisa.hu](http://www.oecd-pisa.hu); [www.timss.hu](http://www.timss.hu); [www.pirls.hu](http://www.pirls.hu); [www.kompetenciameres.hu](http://www.kompetenciameres.hu)).

### *The Impact of IEA on the Development of Assessment Methodology and Data Analysis*

The 40 years of collaboration with IEA and the 10 years with OECD-PISA were instrumental for Hungary in building expertise in educational measurement. Professionals participating in the international surveys have mastered the standard procedures of data development, which have become stricter and more sophisticated along the years. Also, keeping up-to-date in data analysis has been an important con-

dition of participating in the international comparative surveys since each participating country must use the database built and weighted by the international centre. Planning and undertaking independent or secondary data analysis requires some sophistication in both statistical analysis and information technology.

Building and retaining a professional survey team is, therefore, a real challenge. In Hungary we built a team that could turn the lessons learnt in survey methodology into “good practice” applied in national assessments as well. By now standards and procedures of the National Assessment of Basic Competences match international standards in sampling and weighting, developing assessment frameworks and test plans, the application of IRT in test statistics, quality assurance in testing, the development and use of scales, timely reporting and dissemination of results, and in developing data products and analysis tools.

Since the end of the 90s it has been a standard requirement to develop the meta-databases (data thesaurus) and to make public the full technical description of the assessments. The complete data product of each national assessment is now available for further research.

## Conclusion

Hungary has had an established collaboration with IEA for over 40 years. Thanks to the founding fathers of IEA and especially to Torsten Husén and Neville Postlethwaite, and those who succeeded them as Chairs of IEA, this collaboration has always been fruitful and a very stimulating learning opportunity for all those who have had a chance to participate in one or several of the studies.

However, this could not have happened without the professional responsibility, enthusiasm and stamina of the Hungarian researchers and educational policy makers – Árpád Kiss, Zoltán Báthory and Péter Vári, to name the most prominent supporters of IEA – who established

this research contact and helped it survive throughout the Cold War and the political uncertainties of the years following the dissolution of the USSR.

Most important, there has been the commitment and hard work extending over more than 40 years of so many: professionals working on the studies, policy makers willing to stand up for political and financial support for the studies (sometimes more than a little courage was needed for that), and students and teachers who supported IEA by responding positively to the invitation to participate in the studies. The high response rate for all IEA (and PISA) studies in which Hungary participated proves that there has been and is nationwide consensus regarding the value and quality of the studies. One example illustrates the commitment of the teaching force: during a big flood in East Hungary in 1970, one of the sampled schools swamped by the flood, did not forget to rescue the survey material and to send it back completed after the flood subsided. Amidst the tragic events, the students and teachers of a small village school still felt responsible for the success of the survey.

## References

- Báthory Z., Ballér E., Kádárné Fülöp J., Kozma T., Falus I. & Kozák G. (1973). A hazai IEA vizsgálat eredményeiből [A selection of the first results of the IEA study in Hungary]. *Pedagógiai Szemle [Pedagogical Review]* 23, 631-694.
- Báthory, Z. (1992). Hungarian experiences in international student achievement surveys. *Prospects* 22, 434-440.
- Coleman, J. S., Campbell, E. Q., Hobson, C. J., Mcpartland, J., Mood, A. M., Weinfeld, F. D. & York, R. L. (1966). *Equality of educational opportunity*. Washington, DC: US Government Printing Office.
- Foshay, A. W., Thorndike, R. L., Hotyat, F., Pidgeon, D. A., & Walker, D. A. (1962). *Educational Achievement of Thirteen-Year-Olds in Twelve*

- Countries*. Hamburg: UNESCO Institute for Education.
- Husén, T.(Ed.). (1968). *International Study of Achievement in Mathematics. A Comparison of Twelve Countries Vols 1-2*. Stockholm: Almqvist & Wiksell.
- Kádár-Fülöp, J. (1985). The CTD reading study. In: Z. Báthory.& J. Kádár-Fülöp (Eds.), *Educational Evaluation Studies in Hungary, Evaluation in Education: An International Review Series 9*, 117-151.
- Kiss Á., Nagy S., & Szarka J.(Eds.). (1979). *Tanulmányok a neveléstudomány köréből [Studies in Educational Sciences 1975-1976]*. Budapest: Akadémiai Kiadó, pp. 153-275.
- Vári, P. (Ed.). (1999). Monitor '97. A tanulók tudásának változása. [Monitor '97: Changes in Student Achievement]. Budapest: Országos Közoktatási Intézet.
- Vári, P., (Ed.). (1997). Monitor 95: National Assessment of Student Achievement. Budapest: Országos Közoktatási Intézet.

## CHAPTER 19

# **IEA experiences from Latvia, Lithuania, Slovak Republic and Slovenia**

*Hans Pelgrum, Viera Blahova, Rita Dukynaitė,  
Barbara Japelj Pavešič<sup>1</sup> & Andris Kangro*

### **Introduction**

The aim of this chapter is to report on the IEA experiences of several central and eastern European countries, all of which joined this organization in the early 1990s.

This chapter is compiled from the answers to a number of questions to which the various authors replied. First some contextual information about these former communist countries will be presented, followed by a summary of their reasons for joining the IEA. We will then describe how the IEA assisted these countries in creating the research capacity needed for conducting the assessments, followed by an overview of the IEA studies in which these countries participated and a short description of some relevant findings from recent IEA assessments for these countries. This will be followed by a section on what these countries see as the impact of IEA assessments on the field of education - such as policy making, educational interventions and the research community. In the final section the authors reflect on IEA's future and offer suggestions for IEA's further development.

### **Context of Countries in Central and East Europe**

Although IEA was established over 50 years ago, it is only quite recently that many of the former communist countries from central

---

1. There are two other co-authors from Slovenia: Janez Justin and Mitja Čepič Vogrinčič.



and eastern Europe have joined the organization. Hungary - represented by Zoltan Bathory - was the first of these countries to join the IEA in the early 1970s, participating in the Six Subject Survey. It was not until the 1990s that many other former communist countries followed. Among these were Latvia, Lithuania, the Slovak Republic, and Slovenia, four relatively small countries varying in surface area from roughly 20,000 square kilometres (Slovenia) to 49,000 (Slovak Republic) and 65,000 (Latvia and Lithuania). The population sizes vary from 2 million in Slovenia and 2.3 million in Latvia to 3.5 million (Lithuania) and 5.5 million (Slovak Republic). Latvia and Lithuania are located on the Baltic Sea, facing Finland, while Slovenia is at roughly 1400 km distance located on the Mediterranean area. The Slovak Republic (created after the split of Czechoslovakia in 1993) is very centrally located and surrounded by the countries Austria, Czech Republic, Hungary and Poland.

The expansion of IEA's membership in the past two decades was in no small part due to the addition of the central and eastern European countries. These countries, which had been living under an authoritarian socialist regime, became democratic-- a change that was often accompanied by the emergence of new or newly independent nation states. The four countries whose IEA experiences are represented here, all experienced life under the two political systems although their individual backgrounds are quite divergent. In the late 1980s/ early 1990s Latvia and Lithuania regained their pre-World War II independence from the disintegrating Soviet Union; Slovakia had an orderly and peaceful 'divorce' from the Czech Republic; whilst Slovenian independence from Yugoslavia was achieved only after a short military conflict with the federal army. After World War II, and prior to 1990 Latvia and Lithuania were part of the Soviet Union and its economy, whilst as part of Czechoslovakia Slovakia was under the Soviet Union's direct influence within the Warsaw pact. The (Czecho)slovakian economy and trade was also heavily based on the strong links with the Soviet Union. While there was trade with the western coun-

tries, its share was minor. The Slovenian situation was different, as Yugoslavia was developing its own strand of socialism independent from the so-called Eastern Bloc, with its own economic policy heavily based on trade with western economies.

Both political and economic development in the four countries post-independence were also diverse, but one can argue that several processes were common to all. One is the development of a 'capitalist' production system (i.e., the privatisation of state-owned companies and public services), and another is the political integration into the European Union. This latter process of integration with international networks forms the context within which these countries' involvement in the IEA should be understood.

It is also important to summarize how the educational systems in these countries have developed in the past two to three decades.

*Latvia's* school system was substantially reformed during the country's transition from the Soviet authoritarian system to that of an independent and democratic state. The reforms included, among others: centralized examinations and state tests, school accreditation, introduction and improvement of education programs, standards and curricula, improvement of study materials, decentralization of school financing, and international assessment of student achievement.

The improved quality of education in *Latvia* was confirmed in the latest study by McKinsey (Mourshed, Chijioko & Barber, 2010) that revealed Latvia to be one of 20 countries (or rather education systems) included in the study where the improvement of the school system was the greatest. The study also included Lithuania, Poland, Slovenia, England, Singapore, and Korea. The criterion for including Latvia and other countries in this new McKinsey study was the country's improved achievement levels in consecutive rounds of the international studies OECD PISA, IEA TIMSS, IEA PIRLS. Moreover, the increase had to be stable for at least a 5-year period, in several studies and in several subject areas. According to the McKinsey study, Latvia

(but also Lithuania and Slovenia) is one of the countries where the school quality improved from fair to good from 1995 till 2007. The improved achievement levels in these countries are due to the education reforms, which strengthened the education system through a combination of measures such as: collecting important data on student and school achievement; increasing the responsibility of schools and teachers for introducing (ensuring) the results of one's work; developing models for financing and organizational structure; and using the student-centered learning models.

The restoration of statehood in Lithuania has created new possibilities for social, cultural, economic and political development that correspond to national aspirations. Like other post-communist countries, the Lithuanian society is experiencing a fundamental historic shift. This provides a unique opportunity for Lithuania to be part of the community of democratic European nations, and to fully liberate the creative energies which were repressed during the years of occupation, and to create a modern, open, pluralistic, and harmonious society of free citizens. Ideological polarization, dynamic economic and social relations, the increasing importance of information technology and scientific production, and the strengthening of individual independence and initiative are characteristic of this transitional phase. In 1988, Lithuania began to reform its education system according to the underlying principles of national affinity, democracy, humanity and change. At the end of 1992, a proposal was adopted for the overhaul of the entire educational system. These reform measures are currently being implemented in compliance with the provisions of the State Strategy for Education 2003-2012 as approved in 2003 by the Seimas (Parliament) of the Republic of Lithuania. According to this, the principal objective of education is to help every child to understand the world and become an independent, active and responsible person, and to assist every adult in acquiring a qualification and maintaining it by means of lifelong learning. This is to be made possible by the increasing variety of schools: non-state schools are set up alongside state

ones; secular schools exist alongside the schools representing various religious confessions; there are new schools for ethnic minorities; schools that employ alternative educational methods. The legal basis of education has been updated, and the state system now comprises a 12-year general education and two-stage higher education (Bachelor and Master studies). The unified Soviet educational programmes were replaced with a new curriculum providing greater freedom of choice for the individual learner. In 2002, procedures for financing general education schools were changed, and the so-called pupil's basket was introduced. These changes means that rather than financing groups of students, funds are now allocated to individual students. The 'pupil's basket' is a state grant targeted for education. The Lithuanian educational system now focuses on modernisation of learning, improvement of conditions of learning, and harmonisation of the education system. The Ministry of Education and Science now supplies municipalities with yellow school buses so that students living in distant villages can attend schools, and is also providing schools with computers. Since 2001, millions of Litas (about 290,000 Euro) have been allocated each year for renovation of schools, and to date one-third of the country's schools have been renovated. Goals and objectives of the Ministry of Education and Science are: to implement a national system of formal and non-formal education that promotes positive social attitudes towards education and creates conditions for lifelong learning in a changing democratic society. In order to meet its goals, the Ministry develops one-year and long-term educational investment programmes; approves requirements for the regulations of state-run and municipal schools; approves the general curriculum content of formal education, and sets achievement levels; organizes and coordinates the accreditation of the secondary education programme; approves the procedure for learning under general education programmes and the procedure for organization and implementation of the Matura exams (final exam at end of secondary school); and establishes, develops, and reorganizes vocational schools and approves general vocational education plans.

In 2008, the National Council of the *Slovak Republic* approved a new Education Act. The Act mandated that the education and training in schools would comprise a mandatory curriculum ( about 75%) as well as an open-ended curriculum ( 25-30%). The compulsory curriculum is prepared by the state authorities, while the open curriculum or school 'profile' is decided by each school. This represents a significant change in education in the Slovak Republic, as school autonomy is strengthened and each school has the freedom to implement a curriculum that suits its specific orientation (science, environmental,, crosscurricular, or even new subjects, etc.) as well as the needs of the region, employers and other stakeholders. The state curriculum is approved by the Ministry of Education, while the school's chosen curriculum is approved by the school board. In terms of foreign languages this is a compulsory subject from the third grade of primary schools (pupils aged 8) with aim that graduates of upper secondary will have mastered two languages. Starting in September 2011, the English language will be compulsory as the first foreign language. ICT is compulsory starting in the second grade of primary school (pupils aged 7) and is taught throughout primary and lower secondary school as a separate subject. ICT is also taught in vocational schools and computer science is taught in upper secondary school. As stipulated in the Education Act, monitoring and evaluation of quality of education/ training is the responsibility of the National Institute of Certified Measurement in Education and the State Inspection for Education. The expenses of state schools and other educational establishments are covered through public funds from the state budget, municipality budget and/or budget of higher territorial units. All schools (public, private, church) that provide education preparing students for future jobs receive funds from the budgetary chapter of the Ministry of Education. Due to budget cuts on education, the number of schools has decreased, some schools have closed completely while others have united into groups of schools under one administration.

As a republic within federal Yugoslavia, Slovenia was able to exercise

control over its educational system, determining its own educational policy in terms of curricula, textbooks and infrastructure. In fact, Slovenia actually joined IEA in 1989, even before it became independent in 1991. Therefore the new independent status of the country did not result in an abrupt rift in educational practice. In the mid-1990s there was a broad reform of primary and secondary education, aiming at a greater level of autonomy for both schools and teachers in choosing the ways to achieve the educational goals that were still centrally prescribed. The reform was set up after a public debate. With the reform came several types of external assessment of pupil's achievements, such as national assessments of knowledge after grades 3, 6 and 9 (focusing on mathematics and mother tongue), which along with international studies constituted an important tool in monitoring the quality of the educational system.

### **Reasons for Joining the IEA**

The four countries joined IEA in the late 1980s (Slovenia) or early 1990s (Latvia, Lithuania, and the Slovak Republic). The main reason for wanting to participate in international comparative assessments was the concern over the quality of education and the need for international benchmarks in order to be able to make inferences about this quality. This was manifest in different ways in the four countries.

After regaining independence in 1991, *Latvia* was very keen to have international comparisons of education quality as the country had begun a full-scale education reform including the elaboration of educational standards in all school subjects.

In 1988, before *Lithuania* became independent, the country prepared the Concept of Education, which stipulated the implementation of policies for the improvement in the quality of education. In 1990, the country started making efforts to enter the international arena of education. Thus, Dr Darius Kuolys, who was then Minister of Culture and

Education, made contacts with western countries in the quest for information, international research and studies in which Lithuania might be able to participate, as he felt this would offer the country an opportunity to improve the quality of its education. He also felt that the education system would benefit from an external evaluation.

*Slovenia* joined IEA because the Matura exam had been revised and the country felt it necessary to conduct an external evaluation of this and other reforms in secondary education.

From 1990 - 1992, the Minister of Education in the *Slovak Republic* was Jan Pisut, a professor of physics. He understood the value of having the country's students assessed in relation to students in other countries. He believed that math and science were important subjects in Slovak education, so he thought it made sense to be involved in TIMSS. In 1994 when the Civic Education Study was starting up, Slovakia was interested in conducting research on the latest topics in civic education, and wanted to see how their civics curriculum compared in policy, content, and pedagogical practices with other countries. Slovakia was eager for information on new trends in civic education and the improvements going on in other countries. Because the Slovak Republic was beginning to build a new democratic country, involvement in this particular study was highly interesting.

### **Becoming Qualified in Conducting IEA Assessments**

When these countries joined IEA, they had no/very few people qualified to conduct (national) assessments. IEA has assisted these countries in building up staff capabilities. For instance, *Slovenia* developed experience in how to conduct IEA assessments by repeating the Second International Mathematics Study (SIMS) in 1988/89 when SIMS was internationally already finished.

A more systematic professional development project started - with support of the Dutch government - in 1994 and lasted until 1996. It fol-

lowed from the 1993 IEA General Assembly when Zoltan Bathory (GA representative of Hungary) proposed the launch of a training project, for the IEA Network for Central and East Europe (known as IEANCEE). The project was coordinated by Hans Pelgrum, who was later joined by Neville Postlethwaite. It was aimed at building IEA research capacity in nine post-socialist IEA countries – Bulgaria, Czech Republic, Hungary, Latvia, Lithuania, Romania, Russian Federation, Slovak Republic and Slovenia. Young researchers from Austria and Norway also participated in the professional development activities of this project, which resulted in an interesting publication (containing descriptive comparative statistics as well as explanatory analyses) based on TIMSS 1995 data (from schools, teachers and students) about math and science education in the nine central and eastern European countries participating in the IEANCEE project (Vari, 1997).

### **Participation in IEA Studies and Global Trends as Shown in the Data**

Latvia, Lithuania, the Slovak Republic and Slovenia participated (as is shown in Table 1) in several of the recent IEA assessments.

**Table 1.** Participation of Latvia, Lithuania and Slovenia in recent IEA assessments

Country	Primary education				Secondary education	
	PIRLS		TIMSS		TIMSS	
	2001	2006	2003	2007	2003	2007
Latvia	√	√	√	√	√	-
Lithuania	√	√	√	√	√	√
Slovak Republic	√	√	-	√	√	-
Slovenia	√	√	√	√	√	√



It should be noted that all four countries also participate in the OECD-PISA studies.

For illustrative purposes, some observations from the most recent TIMSS reports (Martin et al., 2008; Mullis et al., 2008) are summarized:

- The league tables for mathematics show that all four countries scored above the international average on mathematics at grade 4. This was also true for Lithuania and Slovenia at grade 8 (Latvia and the Slovak Republic did not participate at that grade level).
- The developments over time with regard to mathematics achievement show that in Latvia at *grade 4* there was improvement from 1995 to 2003, while the average score stabilized between 2003 and 2007. In Lithuania this average score was also stable between 2003 and 2007 (Lithuania did not participate at grade 4 level in 1995). In Slovenia a steady improvement took place between 1995, 2003 and 2007.

At *grade 8* level (in which Latvia did not participate in 2007) Lithuania showed an increased achievement level between 1995 and 1999, with the score stabilized since 2003. Slovenia has seen grade 8 achievement improve slightly since 2003.

As evidenced in Table 1, there is no TIMSS data available for the Slovak Republic.

- The gender differences are negligible with regard to average mathematics achievement. In the Slovak Republic and Slovenia, at 4th grade level boys scored slightly (but not substantially) better than girls.
- On attitudes: in 2007, for all four countries at the 4th grade level, the percentage of students with high positive attitudes towards mathematics were around the international average of 72% : Lithuania (74%), Slovenia (71%), slightly lower in the Slovak Republic (68%) and Latvia (65%). However, at the 8th grade, these percentages were much lower, namely 38% in Lithuania and 25% in Slovenia (the international average was 54%).

## Key Actors Driving Participation in the IEA and its Studies

Over the years, there were different persons acting as country representative in the IEA General Assembly, and there were different individuals who played an important role in preparing the application for membership of IEA.

In *Latvia*, the Ministry of Education and Science (Minister Andris Piebalgs, now EU commissioner) supported the University of Latvia to act as a member institution in IEA. The Ministry also supported the establishment of an IEA national research centre at the University of Latvia of which GA member Prof Andris Kangro became the head. The opening ceremony of the IEA national research centre of Latvia (1993) was attended by, amongst others, the IEA Executive Director Bill Loxley, IEA researchers from Sweden, the rector of the University of Latvia, and Ministry officials.

In 1996 the IEA national centre of Latvia evolved into the Institute for Educational Research of the Faculty of Education and Psychology (of which the first director was Prof Juris Zakis, also rector of the University of Latvia; current director of the institute is Prof Andris Kangro). In Latvia the first NRCs were Dr Indra Dedze (Reading Literacy Study), Prof Andris Grinfelds (COMPED), and Prof Andrejs Geske (TIMSS).

At the start of *Lithuania's* participation in IEA, Dr Algirdas Zabulionis from the University of Vilnius was invited and appointed as member of the IEA General Assembly. Later, in 2004, he was replaced by Dr Rita Dukynaitė, deputy head of the Strategic Planning and Analysis Division of the Ministry of Education and Science. The Ministry had chosen to have its own affiliated institutions responsible for international studies and wider educational monitoring and policy analysis. The first of such institutions was the Centre for Informatics and Forecasting (CIF). After the establishment of the National Examination Centre (NEC) in 1996, Dr Zabulionis was appointed the NEC director and the main coordination of international studies was moved there. Dr Zabulionis invited his university students to work in the area of

international research: these included Gediminas Trakas (who was involved in TIMSS), Dr Aistė Mackevičiūtė-Elijio (PIRLS), and Dr Jolita Dudaitė (TIMSS), who later wrote a dissertation based on the international comparative data from IEA. CIF also employed Dr Lina Markauskaitė as national coordinator for COMPED, and at the Institute of Pedagogy Prof Irena Zaleskienė implemented CIVED. In this way, a variety of institutions subordinate to the Ministry of Education and Science have responsibility for international research. Their representatives are appointed as national coordinators in IEA assessments, and an employee of the Ministry of Education is appointed for maintaining the link between policy and research and for coordination of the research on all student achievement (including national projects).

The first representative for the Slovak Republic in the IEA was Dr Jan Herber, director of the Research Institute for Education (Vyskumný ústav pedagogický VÚP), which in 1994 was renamed the State Institute for Education (Štátny pedagogický ústav ŠPÚ). The first Research Coordinator for TIMSS was Dr Rozália Vaculíková (from VÚP), who was succeeded in 1992 by Vladimír Burjan. The studies were conducted by the Štátny pedagogický ústav (since 1994 the National Institute for Education), affiliated with the Ministry of Education. Viera Blahova joined the ŠPÚ in 1995 and was a research assistant to the national coordinator of TIMSS. Her responsibilities were to take care of the communication in English, to prepare materials for printing, to code the short answers and responses to open questions, and to write parts of the national reports. Later she was the national coordinator for SITES M2 and SITES 2006. Since 2004 she has worked as state advisor in the Ministry of Education, Science, Research and Sports of the Slovak Republic.

For Slovenia, Marjan Šetinc was the first general assembly representative, replaced in 2000 by Janez Justin. Barbara Japelj Pavešić has been involved with IEA since 1991 as a SIMS statistician and today she is the national research coordinator for TIMSS.

## Early Experiences with the IEA

The representatives from these countries are quite satisfied with the professional and collegial atmosphere in IEA and attribute part of the IEA success to this.

*Latvia* referred to the warmth and support of IEA colleagues, stemming from the idea of 'IEA as a family' espoused by the (at that time) Chairman of IEA Prof Tjeerd Plomp, and also noted the principle of 'learning by doing' that seemed to be the policy applied to new member countries.

The Lithuanian experience has been that the success of international cooperation depends on the IEA leadership and leaders—their persuasiveness and charisma. Countries differ in many aspects: culture, traditions, economy, history, experience in education, values, etc. In order to sustain interest and see the benefits of research, the professionalism of researchers is important, but not sufficient: tolerance, empathy, and a sincere intellectual and spiritual intuition are required. The IEA leadership has maintained warm interpersonal relations and has gradually included new members in the organization and helped them adapt. The IEA leadership seems to keep up with the new trends in education theory, the expectations of society, and changes in society—which Lithuania believes is essential. Although NRCs of the studies meet only once or twice a year, and primarily to discuss the study and receive training, there is a very strong feeling of a global family, and every meeting is a bit like a family reunion. There is also an openness in the search for solutions, despite the wide diversity of situations/economies/cultures, etc., characterizing the different participating countries. Thus, the IEA provides an excellent example of authentic and caring global cooperation.

In the *Slovak Republic* in the early 1990s, there was very little experience with large-scale assessments and procedures for data collection and data processing. The workshops that were organized in conjunc-

tion with IEA projects, and the printed materials that were made available were of great help to the Slovakian researchers, who ultimately acquired a great deal of knowledge about the theory and logistics of large-scale educational assessments. The financial assistance (from the EU program PHARE and from the World Bank), too, was invaluable for conducting the national assessments (especially memorable was the help of Bob Kozma in locating funds for participating in SITES-M2). In all these endeavours, the IEA colleagues from the Czech Republic provided valuable advice; in turn the Slovak researchers shared their knowledge and experience with colleagues in Romania.

*Slovenia* greatly appreciated the experience and knowledge of key persons in the IEA. Moreover, *Slovenia* found the IEA to be always very helpful and supportive, especially in its assurance of financial help for countries which could never have participated in international studies for lack of sufficient financial resources. *Slovenia* also expressed gratitude for IEA support in terms of training, especially regarding methodology of those studies.

### **Perceptions on how IEA Developed or Changed since the 1990s**

One of the questions addressed to the authors was “In your perception, how has IEA changed since your country joined the association?” Another question was: “What was the general philosophy of IEA at the time you joined IEA? Have you noted any changes in this philosophy, since then?”

Three countries remarked that IEA had changed over the years.

In *Latvia* it was observed that the IEA philosophy and way of working seemed to have changed to some extent. Initially a research-oriented international community (where each country was represented by a research organization), it had become a more centralised international institution for measuring policy-oriented educational indicators using

relatively highly standardized routine procedures (in many countries Ministry officials replaced research representatives in the IEA General Assembly).

The *Lithuanian* experience is that originally the IEA was more focused on students' degree of subject knowledge, whereas now it focuses more on abilities and skills. IEA has expanded; more countries are participating in its activities, which Lithuania feels is a considerable challenge to the IEA. For example, there are more specific aspects and cultural issues that should be measured, assessed, and interpreted, because the world is becoming more varied, there are no countries made up of a single ethnic group, schools are composed of students with different religions, different experience, traditions, and values. On the other hand, assessments of the core area of IEA research must continue on a regular basis, and certain areas must be monitored periodically throughout the years.

*Slovenian* participants observed that today IEA is more open to the suggestions from different countries that are participating in studies, and communication seems to be more frequent and organized. Different studies (e.g., TIMSS, PIRLS, etc.) have become more independently managed (with their own teams) and more academically oriented. In the last few years, academic research appears to have been given more emphasis.

### **The Impact of IEA Assessments on the Field of Education**

In all four countries, it is clear that the IEA assessment studies have influenced education in a variety of ways, including educational policies, school curricula and the research community.

#### ***Educational Policies and Curriculum Reforms***

In *Latvia*, IEA studies influenced education policy and practice for the 19 years that Latvia has participated in IEA activities. For, example, after the COMPED study, researchers were able for the first time to

elaborate standards for the subject 'Informatics'. Further, both the mathematics and science curriculum, as well as the centralized examinations were influenced by the TIMSS studies, and the CIVIC study led to changes in social sciences subjects at the elementary level.

In *Lithuania*, too, educational policies—especially curricular reforms—were influenced by the IEA assessments. For example, the first IEA study in which Lithuania participated (TIMSS 1995) was conducted at a time when the entire educational system was in upheaval. The TIMSS frameworks were very useful in developing new curricula for mathematics and science. Lithuania's participation in TIMSS accelerated changes in the mathematics and natural sciences curricula, as they were amended to be more modern and closer to the TIMSS curriculum framework, and problems similar to the ones addressed in TIMSS tests appeared in teaching aids and examination tasks.

In the *Slovak Republic* there were no educational assessments at all in the early 1990s; yet, today, some 20 years later and after the advent of IEA studies, many assessments are conducted. This has become so important, in fact, that a new national institute targeting assessment has been established: the National Institute for Educational Assessments ("Národný ústav certifikovaných meraní vzdelávania": NÚCEM). The institute is responsible for both national and international assessments and a part of the school leaving exams. IEA has been influential in other ways: certain indicators from IEA studies have been included in policy documents aimed at monitoring educational progress (e.g., the student computer ratio); a number of curriculum changes, too, are due to the country's participation in IEA studies. The studies revealed that some topics were taught earlier in other countries (e.g., basic concepts of statistics in the mathematics curriculum); also it became apparent that Slovak students had difficulty answering open-ended questions, or that they were strong in one topic and weak in others. Participation in the IEA research community was also experienced as very useful, as all National Research Coordinators were actively involved in reviewing the international instruments. There was never a

feeling of being excluded, which ultimately led to cooperative projects with people from IEA studies (e. g., the EU project, Innovative Didactics via web-based learning, coordinated by Hans Pelgrum).

In *Slovenia* it appears that since participation in IEA, attitudes towards international comparative studies have changed – they are now seen as a valuable way of external evaluation of the school system. External evaluation is now in Slovenia's basic educational legislation (e.g., one of the policy goals is to achieve comparative international standards). Participation in IEA studies led to curriculum changes as well (particularly regarding content in some syllabuses and curriculum, especially for mathematics and science). IEA studies have also influenced the design of curricula and the attitudes and views on education in general.

### ***Educational Research***

The authors were asked: "During the time that your country participated in IEA projects, how do you perceive the contributions of IEA to the field of educational research in your country?"

In all countries, the IEANCEE project (IEA Network for Central and East Europe) was experienced as an excellent initiative for building research capacity.

The contribution of IEA studies to educational research in Latvia has been relatively strong. In the Faculty of Education, Psychology and Art at the University of Latvia, IEA studies influenced the content of Master and Doctoral study programmes, whilst many Master and Doctoral theses have used/use data from IEA studies. Some graduates of these study programmes are now working in the Ministry of Education and Sciences in the field of centralized examinations, among others.

In *Lithuania*, IEA's high standards for sampling and other research procedures served as 'lessons' and ideals for Lithuanian national assessments. Also, gradually, study data started to be used. For instance, student achievement data were used in managerial decisions at the nation-



al level--more specifically, the author of the literary texts in primary school textbooks included more informational texts as this was revealed to be needed. Teachers were trained to use a variety of sources, with the aim of improving how they teach children to read and understand texts. Several doctoral dissertations were written that used data from IEA studies, as well as a number of Bachelor and Master theses. Also, in some cases IEA studies are used to teach university students about educational research, using the IEA databases for practicing.

In the *Slovak Republic* educational research is not a high priority, although workshops on the results of TIMSS and PIRLS are still organized.

In *Slovenia* it was observed that the results of the IEA studies over the years have become increasingly recognized as relevant for both the academic field and the general public. Findings from IEA studies are regularly used as a basis for arguing researchers' own theoretical or conceptual constructs; they may also serve as an evaluation of those constructs.

### **The Future of IEA and the Potential for Further Development**

The final two questions asked of the IEANCEE authors were the following: "How do you foresee the future of IEA? What suggestions for the improvement for IEA would you like to share?"

Although individual country perspectives of IEA's future can be distinguished, they all shared the common belief that IEA has a promising future.

*Latvia* sees IEA's future connected with the future direction of educational measurement, assessment of student achievement, evaluation of educational quality – both as a research field and as a measurement field for policy indicators. Latvia sees the future of IEA connected to international organisations, illustrated by the fact that the EU included in the Lisbon strategy for the period 2000-2010-2020 PISA indicators (for achievement in reading, math, sciences and – more specifically -

for low achievers in reading, math and sciences). Actually, from Latvia's perspective IEA collaboration with OECD is considered important and a number of studies can be explicitly mentioned in this context, such as PISA, AHELO-Assessment of Higher Education Learning Outcomes, PIAAC- Programme for the International Assessment of Adult Competencies, TALIS- Teaching and Learning International Survey.

*Lithuania* stresses the continued importance of comparative assessments, observing global trends, learning from other countries, and testing own approaches and trends of reforms--all strong reasons for IEA's future work. Nevertheless the global trend of mobility of students and families is likely to seriously influence interpretation of countries' achievements: while for some time the factor of student origin (the country) will continue to be of interest, this is expected to disappear because migration will have grown affecting all countries of the world. Further, Lithuania feels that IEA should respond more to students with special educational needs in mainstream settings during the procedures of assessment. Moreover, IEA should aim not only to benefit student achievements and economic challenges to education policy, but should also aim to further promote and implement research related to humanistic ideas, the value of the human being, and the development of values.

*Slovenia* strongly supports IEA initiatives to increase the global scope of its research, namely including more Latin American and African countries in future. Greater cultural variety should bring new insights into the interplay of culture, economy and policy on the outputs of educational systems.

In the *Slovak Republic* the IEA studies had a substantial impact on teachers, by showing what was considered important internationally. They were interested in the achievements of their students and adapted their teaching practices on the basis of the findings. It is believed that in the future IEA should help countries understand future trends and how

these can be implemented on a large scale in the education system.

Each country put forward several suggestions for the further development of IEA and its studies, such as:

- *Latvia*: finding a better balance between development of excellent international research and routine measurement of education indicators by strengthening international secondary analyses of IEA research data, including elaboration of recommendations for education policy makers.
- *Lithuania*: in the future, more secondary analyses should be carried out on different topics which are relevant for national policy making. These analyses should result in transparent conclusions and recommendations for education authorities and governments regarding trends and areas of improvements. Lithuanian legislation mandates certain adaptations of the assessment procedures for students with special educational needs (SEN) and IEA should note and include these procedures (for example, extra time during assessment for the blind and visually handicapped pupils, availability of translation to sign language for the deaf, use of ICT for pupils with cerebral palsy, assistants who read aloud instructions to dyslexic pupils, etc.). International assessments should in the future pay attention to these matters.
- *Slovak Republic*: the assessment of 21st century skills is very important and IEA should be a part of the research process leading to regular assessments of these skills.
- *Slovenia*: The academic nature of studies should be more emphasized, as Slovenian educational actors need better understanding of the methodology on which IEA studies are based. Basic concepts and theoretical frameworks should be more explicitly shown and agreed upon among the participants (with more equal influences- not dominant Anglophone/ Anglo-Saxon culture). More European countries should be encouraged to participate, and studies should be extended to include upper grades or even

secondary school level (in Slovenia after 9th grade) (particularly PIRLS and ICCS).

## Conclusions

With regard to the history of IEA in eastern and central Europe, it appears that the four countries interviewed in this chapter had quite similar experiences. Participation in IEA originated from the need for educational evaluation and improvement in a wider context than just one's own country. The IEA and its leadership are appreciated for their professionalism and support to countries that only recently embarked on assessing the achievement of their students and schools. Researchers in the four countries have learned much about the methodology and logistics of educational assessments. And, finally, the expectations for IEA's future are positive, although some suggestions were made for further improvements.

## References

- Martin, M. O., Mullis, I. V. S., & Foy, P. (2008). *TIMSS 2007 international science report: Findings from IEA's trends in international mathematics and science study at the fourth and eighth grades*. Chestnut Hill: TIMSS & PIRLS International Study Center, Boston College.
- Mullis, I. V. S., Martin, M. O., & Foy, P. (2008). *TIMSS 2007 international mathematics report: Findings from IEA's trends in international mathematics and science study at the fourth and eighth grades*. Chestnut Hill: TIMSS & PIRLS International Study Center, Boston College.
- Mourshed M., Chijioko C., Barber M. (2010). *How the world's most improved school systems keep getting better*. Mc Kinsey.
- Vari, P. (1997). (Ed.). *Are we similar in math and science? A study of grade 8 in nine central and eastern European countries*. Budapest: Országos Közoktatási Intézet.



## CHAPTER 20

# **The IEA of 1990-2001: A 50th Year Reflection on Transformation and US Participation**

*Gordon Ambach*

### **Introduction**

In Beijing on a soft September evening in 1990, members of the IEA General Assembly enjoyed an exquisite evening of dining at the Empress's Summer Palace as guests of our gracious hosts of the Chinese National Institute for Educational Research. The General Assembly (GA) met at Xhing Hua University that year, and we were concluding a week of exceptional hospitality, visitation and productive sessions designed to transact IEA annual business and expand personal connections with our hosts regarding participation of the Peoples Republic of China (PRC) in additional IEA studies. After dining, we departed the Palace grounds by strolling, as the Empress and her guests once did, along a lakeside arbor way covered with fragrant flowering jasmine, bidding farewell to the Palace and its extraordinary marble boat which served as a permanent play-station in the middle of the lake.

The event, and the previous week of sessions, was my introduction to the IEA GA. The meeting was also a time to be in Beijing not long after the clashes in Tian'anmen Square. IEA had planned for the meeting well before the Square tragedy. With our hosts urging, our session was held as planned, and turned out to be the first international gathering to convene in Beijing after the extraordinary conflict. IEA met as the city and the PRC prepared to move on, hosting the Asia-Pacific Games for the first time and getting ready to host the next Olympic Games for the first time. Our visit left indelible impressions of Chinese culture and history, contemporary issues, and the intense national interest to participate in comparative international performance, including edu-

cation. The experience was a forerunner of the coming decade for expanded international education studies with increasing numbers of countries across the globe.

I recall this meeting as one bookend of an unexpected, decade-long venture with IEA, an organization with a significant mission and exceptional colleagues dedicated to creating and conducting international comparative studies of education. To be part of IEA during this period of increasing commitment to international studies by policy makers, educators and education researchers was most rewarding. Enhanced government funding, coupled with advances in technical procedures and international communication capacities, substantially changed the work. The changes included transformation of the IEA organizational structure for designing and implementing studies; expansion of types of studies and measures; increase in the number of participating countries; and, for some members, reshaping of organization responsibilities and protocols for participation. During the decade, the people of the Association truly recast IEA to serve greater worldwide appetites for more useful information on comparative student achievement and more illuminating insights on the conditions and processes of learning related to the results. The scope of global education measurement was profoundly changed.

This reflection offers an overview of the IEA transformation and its impact on US participation. The observations are set in the context of broad developments in education occurring in many parts of the world, particularly the reforms of assessment and accountability systems. The context also includes significant geo-political influences of economic, cultural and strategic power competition which stimulated global attention to the relations between education results and economic and security futures. As a story of the 1990s, it is appropriate to begin with the main Association event of 1990 – the GA annual meeting in Beijing just at the time the PRC was beginning to participate in IEA studies. And, to close with another IEA bookend in 2001, the planned GA for Marrakech, Morocco, October 2001, an event never

held as a consequence of the 9/11 attacks on the World Trade Center in New York and the Pentagon in Washington. The potential for travel disruptions and safety concerns led to cancellation, an omen of the coming decade during which transportation and activities around the globe often fell victim to terrorist threats. At the time, the global uncertainties seemed likely to undermine the decade-long trend of escalating interest in international connections and cooperative studies. But this story ends in 2001; the actual record of the first IEA decade of the 21st century is the province of other authors. My term as US representative to the GA closed with my retirement as Executive Director of the Council of Chief State School Officers in July 2001. My venture in the IEA ended as unexpectedly as it had started, with the Marrakech 2001 GA cancelled.

### **Personal Journey with IEA**

To understand my perspective on the transformation of IEA in the 1990s, I offer some background on why my participation in the GA in Beijing, and planned participation in the Marrakech GA, are referred to as unexpected. Although I had known about IEA studies before 1988, I had not participated in developing or implementing them either in the United States or internationally. My designation as the representative to the GA (hereafter referred to as 'GA rep') for the US, and subsequent selection for the Standing Committee, occurred as coincidental to the reorganization of responsibilities through which the US determined how to participate in IEA studies after 1988. This story is recounted as a case study of the way one IEA member reshaped links between the interests and resources of its federal government and the capacities of non-governmental research and education organizations to participate in IEA studies. The objective of the change was to ensure that governmental expectations for supporting the studies were realized while the independence of study directors and researchers was preserved. Most important, the change was a



manifestation of how essential the work of IEA had become to the US in providing desired information from international assessments which had strong quality, reliable periodicity, more effective reporting and expanded country participation.

### **US Participation in IEA Studies**

A pivotal point for US interest and investment in international assessments occurred in 1989 at the Charlottesville, Virginia 'Summit' meeting convened by President George H.W. Bush and the nation's governors, at the time chaired by Governor Bill Clinton of Arkansas. This 'Summit Meeting on Education' produced a bipartisan consensus on a statement of National Goals for Education intended to spur action at the national, state and local levels to improve the performance of the elementary and secondary schools and their students. Among the several goals (which were not incorporated in any legislation at the time), the most significant was the first, a statement of expectations for US student achievement levels by the year 2000 in the academic subjects of English literacy, history and civics, mathematics and science, which would exceed the comparative performance of other countries. This bold, optimistic expression signaled a strong belief among these leaders in the importance of education and in the necessity for intense improvement in the quality of our schools which required substantial new efforts at all levels of the education system.

At that time in the US there were additional widespread activities for development of large-scale assessment practices. Many states were authorizing developments of statewide content standards and initiating state assessments. Some states, such as New York which had a history of more than a century of secondary school state exams and had substantially revised its standards and tests beginning in the late 70s, and others, were well on the way to new systems by the late 1980s. Other states were in early stages of accountability reforms. At the national level the National Assessment of Education Progress (NAEP),

originally designed in 1968 as a sampling measure for long-term achievement trends which could be reported only at the national and regional, or multi-state, levels were being re-tooled to enable state-by-state reporting of results.

In anticipation of greater use of international comparative results, in 1988 and 1989, the United States Education Department (USED), through the National Center for Education Statistics (NCES), and the National Science Foundation (NSF) commissioned the National Academy of Sciences-National Research Council (NAS-NRC) to establish a Board on International Comparative Studies of Education (BICSE) to advise the government on participation in such studies—their importance, quality, auspices, voids, and on how to orchestrate US participation. At that time, the US was, of course, already participating in IEA studies through representation in the GA and through research leaders of some international coordinating centers and for US study centers. Participants from the US, together with colleagues from other countries, had helped to build the foundation for the transformation to come in the 1990s. Creation of the BICSE was informed by the earlier experience with IEA and the advice of several persons who had participated in IEA governance and studies.

As a part of the National Resource Council, BICSE was part of a long-standing NAS-NRC function to conduct analysis and provide advice to the US government on major policy and administrative issues in fields such as health, climate, nuclear energy, social welfare and education. The BICSE included specialists in assessment, statistical surveys, international comparative studies, school improvement strategies and education policy making. As Executive Director for the Council of Chief State School Officers, with responsibility for representing the states before the Administration and the Congress on issues of federal education legislation and policy, I was appointed a member of the BICSE.

In the initial design for the Board, the Chair was also to serve as the US

Representative to the IEA GA. The incumbent would represent the US over a fixed term with support services provided through BICSE staff and be guided by the experts on the Board on issues related to the IEA studies. The Chair would designate US study centers and project directors for participation in international studies; serve as the GA rep unencumbered by any other IEA role, for example, as an international center coordinator or US study director; and, maintain continuing contact with USED and NSF officials, as representatives of the agencies funding BICSE and the agencies most likely to fund IEA studies, or have major interests in US participation in international studies.

There was an important context for establishing the BICSE Chair as the GA representative. During this period, there were several developments with regard to establishing and using state and national tests. At the national level, The National Assessment of Education Progress (NAEP) was being reshaped to measure state-by-state results in addition to nationwide results. Progress on the National Goals for key subject fields was to be measured in terms of comparative international achievement results. For these two projects, and others, the issues of which US curriculum standards (frameworks, or content specifications) should be used to help guide development of these assessments, and who should have responsibility to determine them, were heating up rapidly among educators, school boards, administrators, and legislators at all levels of government. Any role for federal agencies, and the federal government, in the preparation of subject content specifications and developing test items for national and international assessments was strongly contested.

The US Constitution is silent on responsibility for education; the authority resides with the states (or, as the state delegate, with the localities). Federal statutes explicitly prohibit the US Education Department from establishing or interfering with the school curriculum. But, to test US students reliably requires the tests to be based on expectations – questions and exercises – of student knowledge and skills in the curriculum content assessed. If federal agencies were to

control the decisions on the questions and exercises, they would, in effect, be establishing the curriculum.

These circumstances required special arrangements to create space between federal funding agencies, and their officers, and persons or organizations designated to represent the country in processes of setting assessment standards, or frameworks, and developing the assessments for the studies. This had been done on the domestic side by having NAEP administered by an independent Board. The frameworks for NAEP assessments in all subjects from 1987 until at least 2001 were all developed through the Council of Chief State School Officers (CCSSO) – the organization which represented all state education agencies and had extensive experience in assisting states with their own standards and test systems. So, the BICSE was created for the international side of assessments, to be the buffer for independence from ‘federal’ control of test construction and administration and to ensure the assessments were not measuring a USED or federal government curriculum.

BICSE started operating in 1988; however, before the Chair began serving as the GA rep, a new issue arose that caused reexamination of whether the Chair could serve in the role of GA rep and also carry certain other responsibilities as Board Chair. As IEA was moving ahead with plans for TIMSS (Third International Math and Science Study, later to be known as Trends in Math and Science Study), the Education Testing Service (ETS), which administered the NAEP program for national domestic assessment, was developing an international counterpart, the International Assessment of Education Progress (IAEP). ETS advanced the IAEP assessment concept as an alternative to the IEA study and sought support both within the US and from other countries. Since one of BICSE’s responsibilities was to advise the USED and the NSF on which assessment, if there were alternatives, should be funded, the circumstances created a dilemma for responsibilities of the Chair. To lead the Committee in comparative analysis of assessment systems, including those of IEA, in order to advise the USED and NSF, while simultaneously serving as the US rep to the GA,

created the appearance of, and, potentially the reality of, a conflict of interest in this case. (And, likely in similar circumstances which might occur in analyzing competing proposals for other studies.) The BICSE founders decided to separate the two roles of the Chair (and the Board), by determining to select another member of the Board to be designated GA rep with full authority to carry out the US IEA responsibilities, including selection of US study directors and centers for participation in IEA studies. The member would continue on BICSE but be recused from any deliberation regarding analysis of qualifications for any international studies in which the US might participate. Neither I nor my organization, CCSSO, sought this designation nor expected it. We were honored to be offered the responsibility, and to accept, because we believed strongly in the importance of participating in IEA studies.

The BICSE made the designation of the Council of Chief State School Officers for five reasons. First, CCSSO supported US participation in IEA primarily for what might be learned from the natural laboratory of different country practices and conditions of education and the relationship of learning strategies to differential student performance results. The states wanted to be better informed about comparative trends in student achievement and to learn more about pedagogical practices and conditions for education in countries related to them. States were interested not only in the nationwide results but hoped that eventually individual state samples and reports could be drawn from the international assessments. The Council had a perspective on overall education policy, particularly centered on school improvement strategies, and the inter-relationships between the federal and state levels in pursuing these strategies.

Second, CCSSO had extensive experience and capacity for organizing the states in developing frameworks for all the subjects assessed under NAEP. This work, which started in 1987, had demonstrated to the federal agencies that the Council effectively developed procedures for determining consensus among the states on this most sensitive aspect of national assessment development—linking testing to what is taught.

And, further, the Council had demonstrated to the states the effectiveness of using this process to ensure their interests were met. The experience with this consensus building across several different content fields, done essentially through state education assessment directors, provided a basis for BICSE to determine that the Council was in the best position to organize the states' perspective in advising IEA study directors, both national and international, on the fit between state content expectations and assessment procedures and those developed by the IEA.

Third, in the US, states were essential partners in helping to draw national samples for IEA assessments and working with local districts and schools to implement testing and gather system data. The CCSSO network of state assessment directors was central to efficient orchestration of that process. Fourth, CCSSO had extensive connections with the US research and assessment community at the core of international studies. This was important for designating the national study directors for international studies and securing advice on IEA studies in addition to that offered through BICSE. And, fifth, the organizational connections provided the means through which federal agency representatives together with representatives of state agencies and nongovernmental research organizations were able to join forces to design, fund and implement IEA studies for the US.

Although the initial BICSE designation was for a two-year trial period, the assignment was periodically renewed, and thus was continuous through the decade until 2001. The period turned out to be stimulating and important to the growth of international dimensions of education, and a privilege for me as both a GA rep and member of the Standing Committee (SC) to share in the work with gifted colleagues who guided the rise of IEA to the center of a new global emphasis on the value of comparative education studies.

### **Three Major Aspects of IEA Transformation**

Through my GA and SC experience I witnessed the emergence of three

ways in which IEA progressed through the decade to be a significantly different organization: the addition of 24 new members from regions across the world over the the decade; the use of new methods to record conditions and practices of learning together with the expanded diversity of educational objectives (besides study of the traditional subjects of mathematics, sciences, reading literacy, and civics); and the development of the necessary IEA administrative and organizational capacity to manage increased funding and study complexity, and to coordinate the essential assessment capacities outsourced to agencies across the globe.

### *Increase of IEA Members*

The increase in IEA members during the 1990s occurred in six regions of the world. The new members, grouped by region, included: **Africa:** Botswana, Kenya, Morocco, and South Africa; **The Americas:** Brazil and Mexico; **Asia:** Chinese Taipei (Taiwan) and Philippines; **Middle East:** Iran, Kuwait, and Turkey; **Europe—Former USSR sphere of influence:** Bulgaria, Czech Republic, Latvia, Lithuania, Macedonia, Rumania, Russian Federation, Slovak Republic, and Slovenia; **Europe-Remaining:** Cyprus, Iceland, Ireland and Scotland.

The new members each had particular reasons for joining. For some, the reason was interest in a particular study; for others it was an interest in developing their own country capacity for implementing assessment systems. For all, the key motive was to learn about their comparative conditions of education related to their performance, either worldwide or regionally. In many cases new members were encouraged to participate by international organizations such as the United Nations, UNESCO and the World Bank and regional organizations such as Asia Pacific Economic Cooperation (APEC). In the aggregate, the new membership added significantly to the range of countries studied and brought greater diversity of education traditions and practices for comparative practices. The trend was set for even further expansion of globally distributed membership in the new century.

### *Expansion of Methods and Types of Studies*

The traditional means of data collection for pedagogical practices, scope of student achievement assessed, and educational domain studied by IEA were significantly supplemented during the decade. Here are three examples. (A fourth proposed innovative study of teacher education and development was in the planning stage for the 2001 GA.)

#### **TIMSS special study of classroom videotaping**

Traditional methods of collecting information about instruction in large-scale assessments have included written descriptions of teacher activities, analysis of lesson plans, written descriptions of time on various tasks, descriptions of classroom environments, and other written or audio records. The major TIMSS study included a supplemental trial study of classroom instruction using videotaping of teaching and student activity. The objective was straightforward—to establish country samples of an audio-video record of pedagogical practices. This was the first time the technology was used in a large-scale international comparative study. The trial put down a significant early marker for IEA in the use of instructional videotaping and the development of protocols for the taping; rubrics for analysis; and directions for record storage to enable analysis of practice compared across classrooms and future study comparisons. The practice of such videotaping is now commonplace in many countries for self-evaluation and improvement of teacher practice, evaluation by supervisors and mentors of teachers, and for teacher portfolios to demonstrate instructional proficiency as part of position applications and/or award considerations. IEA innovation here was early and influential.

#### **Computers and the use of instructional technology in education**

Traditional IEA studies focused on the objectives of student achievement in particular subject fields and the collection of data related to aggregate impacts of resources and methods intended to effect results as measured in that subject. In the latter part of the 1980s, IEA began a



series of studies on the use of computers and instructional technologies which broke out of that design. Beginning with the COMPED (Computers in Education) study (data collection 1989 and 1992), and followed by the Second Instructional Technology in Education Study (SITES), IEA implemented this ground-breaking design. The objective was to help determine use and effectiveness of these technologies during a time of rapid innovation when the desktop computer was becoming available and affordable for classroom instruction. Rather than centering on achievement in a subject, the studies focused on assessing student capability in using computers to accomplish certain tasks. The studies were also designed to estimate the availability of computers in schools and the status of student use.

These early international studies occurred at a period when there were significant controversies about allocation of resources for instructional technologies vs. greater investment in teachers. Those advocating for investment in computers for instruction had an uphill struggle. Today, with the hindsight of more than a decade of incredible advancement in computing and digital technologies, we are astonished to recognize how 'primitive' the concepts measured and the instrumentation were in those early studies. But, at the time, IEA was the bold leader to launch international assessment of the early stages of computer use in schools – now, arguably, the most powerful transformational force in education. This work was a major innovation in education assessment.

### **The Pre-Primary Project (PPP)**

A third example of the expanded repertoire of IEA studies which attained an important stage of accomplishment in the 1990s was the Pre-Primary Project. This study was first envisioned by experts in early childhood development in 1979 and advanced to the IEA GA in 1982 in hopes that the Association would conduct the study. The IEA was chosen because of its reputation and capacity to carry out cross-country studies, organize and prepare researchers to study learning in several countries, and cope with the special complexities of designing

a study of varied learning settings within and across countries – in and outside of schools – for children of this age. The broad goal of the study was to determine the various prevailing aspects of early childhood education called ‘quality of life’ in the different countries and to determine what should be provided for these children to enable them to do better in school following the pre-primary experience.

In 1984, the IEA GA approved the study and designated the High Scope Educational Research Foundation as the International Coordinating Center. The study was designed to comprise three phases to accommodate the complexity and novelty of the approach and resource availability in various countries. Considerable effort was required of each participating country to develop a composite description of early childhood experiences which might include several settings – child care, nursery school, etc. Data collection for this study was through expert observations and through video recording.

During the 1990s, IEA GA members were provided opportunities to review PPP annual progress reports and gain greater exposure to pre-primary learning, a level of education which was not included in other Association studies. This opened access to developments in practices and conditions of learning which were becoming increasingly important in several countries during the 1990s. A significant study accomplishment of this period was completion of a video project which portrayed typical practices of pre-primary learning in 12 countries. The IEA earned significant recognition and respect for including this overall PPP study in its portfolio of projects.

One anecdote about recognition of the work and the High Scope leadership is illustrative. The release of results and the video presentations from the 12 countries received extensive publicity in the US. As part of the release, CCSSO held a special public presentation in Washington, DC, which was attended by representatives of the embassies of participating countries. Each representative was given a copy of their country’s tapes. In a follow-up, we learned that all the embassy welcome

centers were including the video and other study materials as part of their presentation of early childhood life in their country. This was a happy, if unanticipated, consequence of an IEA study; this particular IEA contribution to educational development in the 1990s was clearly considered exceptionally important in many countries.

### *Transformation of IEA Organization and Operational Capacity*

A third major strand of transformation included strengthening IEA's organizational structure and administrative capacity, and creating new consortia of research, data collection and analysis organizations to conduct the studies. These developments helped to enable expansion of IEA membership and increase the types of studies and assessment methodologies, as noted above. The driving force underpinning them, however, was the need to build IEA capacity for implementation of the core IEA studies. These studies were enlarged in scope and complexity of assessments and data collection. They had intensified study schedule demands and they required more effective management of substantially increased resources. Every effort had to be escalated to meet the new interests and expectations of policy makers and sponsoring governments for better and more complete information about the performance of their and other education systems in the studies of mathematics and sciences (TIMSS and TIMSS-R), reading literacy (PIRLS), and civics education.

In virtually every aspect of IEA work, the pressure to perform more effectively, more responsively, on time and with the involvement of many more personnel and organizations drove change. The changes did not occur according to one grand plan or all at once; they were spread over the decade, growing organically as IEA became ready. They are summarized as follows:

*Revision and Clarification of the Member Guidance for Participation and the IEA Bylaws for Governance and Operation.* The revisions of these documents required extensive SC and administrative effort. They were especially important for new members to understand obligations and

responsibilities of membership and to clarify roles assigned to the organizational units of IEA—the GA, SC and other established committees, and the International Coordinating and Country Study Centers.

*Strengthening the Working Capacity of the IEA Governing Structure.* Expectations for the work to be accomplished by the GA and the Committees—Standing (SC), Technical Executive Group (TEG), and Publications and Editorial (PEC)—increased substantially with the scale, cost of studies, and production demands. The Standing Committee was the driving force behind the changes which included: better preparation for the meetings and deliberations of the GA on the proposed and continuing studies; more effective appointments of personnel for the Committees and support for disciplined Committee reviews across studies to ensure consistency of protocols and practices among the studies; greater care in regional representation and proficiencies of appointees to the SC; increased thoughtfulness in selection of GA and SC meeting locations, particularly related to encouragement of new IEA members; and, extensive deliberation and effort on the decision to establish a ‘home’ location for the organization and to establish an IEA Secretariat with a full-time Executive Director and key associates to handle finance and special membership issues.

*Establishment of a permanent IEA Secretariat.* Growth in the scope, size and financing of IEA studies outgrew the pattern of expecting that the Chair of the IEA and the GA could handle both the responsibilities of presiding Chair and those of the chief executive officer. Members of the Association, and the IEA sponsoring organization, the Dutch Ministry of Education and Science, recognized that a division of responsibilities was needed. Fortunately, resources were made available to establish the Headquarters in The Hague (later moved to Amsterdam) and establish the new Executive Director position. This was essential to manage staff and the complex array of contractual arrangements to conduct the studies; maintain contacts with sponsors and with potential collaborative partners in international organizations; manage study oversight of the international coordinators and

relations with country GA reps and country study directors; and, oversee plans for GA, SC and Committee meetings and agendas.

*Development of Continuing International Study Centers and Consortia for Large-Scale Data Collection and Analysis.* Two additional operational actions were taken to advance IEA capacity (1) for conducting large-scale studies with increased participants and regular cycles, and (2) to engage technical organizations which had been established primarily for country, or regional services, to be part of a consortia for providing multi-agency capacity in data collection, processing, analysis and reporting. An example of the first was establishment of the TIMSS International Study Center (later renamed TIMSS and PIRLS International Study Center) at the Center for Testing, Evaluation and Educational Policy at Boston College with the expectation of continuing leadership of studies at four-year intervals for TIMSS and five-year intervals for PIRLS. An example of the second was creation of the consortia including ACER of Australia, Statistics Canada, the IEA Data Processing Centre Hamburg and Boston College for processing the extensive test and data collection for studies such as TIMSS.

*Comment.* These several organizational and administrative changes were not particularly innovative or exceptional in themselves, especially if considered in the context of transformations underway in other nonprofit or commercial organizations at the time. They were, however, sensible, timely and effective, and they were important advances taken by this Association. In 1990, the IEA was essentially a voluntary group of representatives from research institutions or government agencies that had operated for three decades through self-generated agreements to conduct studies. It struggled constantly to generate support for its work and interest in its results. In the decade of the 1990s, the Association matched the expectations assigned to it for producing information on international education conditions and results with a transformed capacity to provide them.

## Value of IEA to the United States

This reflection referred earlier to the 1989 expression by US national leaders that progress on “National Goals for Education” should be measured by comparisons with achievement results of other nations. That expression raised the stakes for deliberations on several activities already underway in our country that centered on developing more explicit expectations, standards and assessments for education at the local, state and national levels. The challenge of expanding participation in international studies substantially added to the complexity of linking new study requirements with assessment reforms being planned at these other levels. We needed to merge systems across levels through an efficient and effective solution. Fortunately, the solution was in reliance on IEA. The US had, of course, been a member of IEA and participated in studies for decades. However, during the 1990s participation in IEA increased immeasurably in value, as illustrated by these points:

1. The US was able to select IEA for the major work in several subject fields and, thereby, expeditiously undertake extensive planning and preparation for the international studies through the extant IEA structure. The partnership remained strong throughout the decade.
2. IEA had established credibility, and maintained credibility, for its processes of developing frameworks and specifications for the subject content to be measured across countries; for sampling; for conduct of assessments; and, for reporting of results. IEA’s credibility, in these respects, was accorded bipartisan support during this period through three national administrations and by state and local education and government authorities across the nation.
3. IEA studies focus on the practices of education and conditions of education as well as comparisons of test results and lead tables. The tradition of contributing insights on how characteristics of

practice are related to student results is part of the core of IEA studies. The commitment to research that informed countries about comparative practices that might improve student performance was an essential value.

4. The design and implementation of regular cycles of studies in mathematics, sciences and reading literacy enabled establishment of trend lines of student achievement related to national goals and helped to relate international to national trend lines. In addition, the availability of special sampling for individual states, and, subsequently major cities, enabled them to participate separately in the studies. This feature has provided a very desirable option for those jurisdictions to compare performance directly with international results.
5. The developments of new domains for study – notably pre-primary learning and use of computers – have broken new ground for both ‘subjects, which have become extremely important growth areas of education. Developments of new methodologies – especially videotaping classroom instruction in the TIMSS and pre-primary studies – provided early applications of using visual and audio recording for evaluating and comparing pedagogical practices. These methodologies have informed subsequent assessment practices and become widely imbedded in professional development and evaluation techniques.
6. The expansion of IEA members has led to insights on education practices and results in regions and countries about which US educational policy makers and practitioners generally had too little knowledge. The IEA studies of this period have provided better understanding of education commitment, pedagogical strategies and impressive student results, particularly in Asia and eastern Europe. They have brought insights on motivations imbedded in cultural traditions and economic ambitions related to enhanced learning opportunities. The media attention given

to IEA study results has sparked US public understanding of the importance of education throughout the world and the value of our comparative results.

- 7 In addition to the public information impact of study results due to expansion of IEA studies and members, there were important personal benefits for those of us who were part of IEA. The exceptional opportunity to work with colleagues across a more diverse range of countries and regions has developed bonds of mutual interest and admiration for new partners in this global research.
- 8 The development of international comparative studies in the 1990s established a resolute US expectation that future measurement of US conditions of learning and elementary and secondary student achievement must continue to be done in an international context.

### **Personal Appreciation for an Exceptional IEA Journey**

The opportunity to serve in IEA was an unexpected privilege and rewarding experience—both personally and professionally. My day job centered on education policy development; I was a **consumer** of research on the results of education. Participation in IEA added a new dimension—working together with experts who created the studies and **supplied** the results. I developed a great admiration and respect for the colleagues with whom I served and for those pioneers who created and developed IEA in the early decades. To me the heart of IEA's contribution is the comparative analysis of the pedagogical practice and conditions of learning related to student achievement in the participating countries. In terms used regularly in the 1990s, the contribution is in describing and understanding the '**opportunity to learn**' which causes the achievement. I highlight this priority with full realization that most public attention to IEA results is on the lead tables and rank order of participants.



There would likely be no IEA, nor IEA studies, if there were no comparative achievement results to report. The roots of the Association, however, are imbedded deeply in learning about what relates to, or accounts for, the comparative results. These are the findings which best help educators, policy makers and the public to invest in the strategies most likely to improve the conditions of learning and student performance. My central purpose in serving with the Association has been to help preserve and strengthen that core of the mission so IEA would advance the capacity into the next century.

I was honored to represent the United States in the General Assembly, to serve on the Standing Committee and to be elected Honorary Member, during a period of exceptional transformation for international studies and for IEA.

**Thanks:** This reflection focuses on the broad accomplishments and the collective contributions of IEA personnel: it offers an overview of the decade, rather than the details of particular projects or of individual efforts. To thank all colleagues individually and recognize their unique contributions would require doubling of this text. May I thank them all by group of affiliation for their generous help with my education in the work of IEA and support for my participation. US colleagues: Board members and staff of BICSE and the NAS-NRC; US Education Department and the National Science Foundation officials; Study Directors for US and International Center Coordinators; and, CCSSO Board members and associates. Among these groups, five persons are especially cited: Dorothy Guilford and Emerson Elliott, USED National Center for Educational Statistics, who were so pivotal in the formation of BICSE and reshaping US participation in IEA; US Secretaries of Education, Cavasos and Riley, for their support of international studies and the IEA; and Ramsey Selden, Director, CCSSO Assessment Center, who guided all aspects of our work with IEA studies. **IEA colleagues:** General Assembly Members, with a special nod to those who served on the SC; International Center Coordinators and Committee Members; and IEA Staff. Among these groups, three per-

sons who served IEA in leadership roles throughout the decade are especially cited: Tjeerd Plomp, my close colleague through all these years, who guided the transformation as IEA Chair; Alejandro Tiana Ferrer, IEA Chair as successor to Tjeerd; and, Hans Wagemaker, GA and SC colleague from New Zealand, and IEA Executive Director.

I close with a special thanks to Constantinos Papanastasiou and Tjeerd Plomp for their initiative with this 50th Year IEA report and overseeing its successful completion. In preparing my piece I have enjoyed happy recollections of them and colleagues and places I would never have known, had it not been for my participation in the IEA. To all of you, I extend much appreciation for the journey from Beijing in 1990 to Marrakech (almost) in 2001!!







# IEA 1958-2008: 50 Years of Experiences and Memories



*Volume 2*



# IEA 1958-2008: 50 Years of Experiences and Memories

EDITED BY:  
Constantinos Papanastasiou  
Tjeerd Plomp  
Elena C. Papanastasiou

*Volume 2*



The International Association for the  
Evaluation of Educational Achievement  
(IEA)



Copyright © 2011 The International Association for the Evaluation of Educational Achievement. All rights reserved. No part of this book may be reproduced in any form, by any means, including, but not limited to, the process of scanning and digitization, or stored in a database or retrieval system, without the prior written permission of the IEA.

International Association for the  
Evaluation of Educational Achievement  
Stichting IEA Secretariaat Nederland  
Amsterdam, The Netherlands  
Herengracht 487, 1017 BT Amsterdam, The Netherlands  
Phone +31 (0) 20 625 3625  
Fax +31 (0) 20 420 7136  
Email [secretariat@IEA.nl](mailto:secretariat@IEA.nl)  
[www.iea.nl](http://www.iea.nl)

Edited by  
Constantinos Papanastasiou  
Tjeerd Plomp  
Elena C. Papanastasiou

Published originally by  
Cultural Center of the Kykkos Monastery  
PO Box 28192, 2091 Nicosia, Cyprus.

Typeset: Maria Nicolaou-Prodrörou  
Printer: Kailas Printers and Lithographers Ltd  
Cover Design: Stelios Smyrillis

ISBN 978-9963-645-78-7

IEA, the International Association for the Evaluation of Educational Achievement, conducts worldwide comparative studies of educational achievement. It is well known for its cycles of studies, such as the Trends in Mathematics and Science Studies (TIMSS) and the Progress in Reading Literacy Studies (PIRLS), as well as studies focused on particular subjects and themes such as civics education and ICT in education.

IEA's founding meeting took place in 1958 at the Unesco Office in Hamburg (Germany). This first meeting, it turned out, heralded the beginning of many impressive international comparative assessment studies. Over the 50 years since that initial meeting, the IEA has developed into a highly renowned research organization with a membership of more than 60 education systems (or countries) worldwide.

In 2008, the IEA celebrated its 50th anniversary at the 49th General Assembly meeting in Berlin where, during a special Roundtable Session, a number of honorary members shared their past experiences in IEA with the IEA representatives present. It was clear that the IEA had evolved over the years to become not only a vibrant, global organization, but also an international community of friends and colleagues, all of whom were deeply committed to their involvement in IEA.

This book presents a mosaic of experiences and memories of IEA members from all over the world, reflecting a wide variety of foci on the IEA and its studies, including: personal memories; how studies were originated, designed and developed; the impact of IEA studies on a country's policy and practice; and perspectives of observers involved in IEA's activities.

The 29 chapters of this book present the reader with a rich collection of memories and experiences. They reveal how relevant and important the IEA has been for educational research, policies and practices, but they also demonstrate how the IEA has, over the past 50 years, developed into a 'global community' of professionals and friends.



# Contents

## *Volume 1*

### **PART A:**

#### **EARLY DEVELOPMENTS - HISTORY AND STUDIES**

1. IEA - From the Beginning in 1958 to 1990  
*John P. Keeves* .....3
2. IEA from 1988 – 1994: An Organization in Transition  
*Tjeerd Plomp* .....41
3. The Second International Mathematics Study (SIMS):  
Intention, Implementation, Attainment  
*Kenneth J. Travers* .....73
4. The IEA Studies in Civic Education: Their History  
from the Late 1960s to 2005  
*Judith Torney-Purta, John Schwille* .....97
5. The International Study of Writing  
*Sauli Takala* .....115
6. IEA Assessments of Information and Communication  
Technologies (ICT) in Education  
*Willem J. Pelgrum, Tjeerd Plomp* .....137
7. The Growth and Development of the Preprimary Project  
*Jeanne E. Montie* .....165
8. The International Civic and Citizenship Education Study (ICCS)  
*Wolfram Schulz, Julian Fraillon, John Ainley* .....187
9. IEA International Research Conferences (IRC) from 2004 to 2008  
*Constantinos Papanastasiou, Elena C. Papanastasiou* .....201

**PART B:****THE CONTRIBUTION OF IEA TO EDUCATIONAL RESEARCH**

10. The Relationship of IEA to some Developments  
in Educational Research Methodology and Measurement  
during the Years from 1962 to 1992  
*John P. Keeves, Petra Lietz* .....217
11. IEA: International Studies, Impact and Transition  
*Hans Wagemaker* .....253
12. The Contribution of IEA to Research and Education  
*Richard M. Wolf*.....273

**PART C:****REFLECTIONS ON CONTRIBUTIONS OF IEA RESEARCH  
FROM THE PERSPECTIVE OF EDUCATIONAL SYSTEMS**

13. The Involvement of African Countries  
in the IEA Studies over 50 Years  
*Sarah Howie* .....289
14. The Contribution of IEA Research Studies to Australian Education  
*John Ainley, John Keeves, Jan Lokan, Petra Lietz,  
Geoff Masters, Sue Thomson* .....317
15. Chilean Participation in IEA Studies  
*Leonor Cariola, Catalina Covacevich,  
Johanna Gubler, Ema Lagos, Marcela Ortiz* .....373
16. The Significance of IEA Studies for Education in East Asia  
*Frederick K.S. Leung* .....389
17. The Impact of IEA on Educational Policy Making in Germany  
*Rainer Lehmann* .....411
18. How IEA Influenced the Education System in Hungary  
*Sándor Brassófi, Judith Kádár-Fülöp* .....431
19. IEA Experiences from Latvia, Lithuania,  
Slovak Republic and Slovenia  
*Hans Pelgrum, Viera Blahova, Rita Dukynaitė,  
Barbara Japelj, Andris Kangro* .....447

20. The IEA of 1990-2001: a 50th Year Reflection  
on Transformation and US Participation  
*Gordon Ambach* .....469

## *Volume 2*

### **PART D:**

#### **MEMORIES OF KEY PERSONS IN THE IEA**

21. Roundtable at the 49th General Assembly, Berlin 2008  
Christiane Brusselman- Dehairs, Tjeerd Plomp,  
Neville Postlethwaite, David Robitaille,  
Judith Torney- Purta, Alejandro Tiana, Tom Loveless  
*Edited by Judith Torney- Purta* .....493
22. The Scientific Contributions of Torsten Husén  
and Neville Postlethwaite to the Development of International  
Comparative Research on Educational Achievement  
*Rainer Lehmann* .....515
23. The Evolution of the IEA: A Memoir  
*Alan C. Purves* .....531
24. Experiencing the World as an Educational Laboratory  
*Alejandro Tiana Ferrer* .....557
25. 50 Years of IEA: Some Reflections about its History  
*Judith Torney-Purta* .....577
26. 50 years of IEA: A Personal Account  
*Kimmo Leimu* .....591
27. Experiencing Innovation and Capacity Building  
in IEA Research - A Memoir, 1963-2008  
*John Schuille* .....627

### **PART E:**

#### **MEMORIES OF FRIENDS OF THE IEA**

28. Reflections on IEA from the Perspective of a World Bank Official  
*Marlaine E. Lockheed* .....711

29. Guiding International Comparative Studies (1981 to 2009): Reflections of a U.S. Program Director <i>Larry E. Suter</i> .....	721
<b>APPENDIX:</b> Author Biographies.....	745



**PART D:  
MEMORIES OF KEY PERSONS  
IN THE IEA**





## CHAPTER 21

### Roundtable at the 49th General Assembly

*Judith Torny-Purta, Editor*

*Participants:* Christiane Brusselman- Dehairs (Belgium), Judith Torny- Purta (USA), Tjeerd Plomp (The Netherlands), Neville Postlethwaite (England), David Robitaille (Canada), and Alejandro Tiara (Spain)

*Moderator:* Tom Loveless (USA)

The 49th IEA General Assembly in Berlin October 6-8 2008 marked the 50th anniversary of IEA as an organization. To commemorate and celebrate, this Round Table of Honorary Members of IEA was organized on Tuesday October 7th 2008. Honorary Members are individuals recognized by the organization for their long-term contributions to the program of international achievement studies (separate from General Assembly members who represent specific countries). The distinguished Christiane Brusselman- Dehairs from Belgium, Tjeerd Plomp from the Netherlands, Neville Postlethwaite from England, David Robitaille from Canada, Judith Torny- Purta from the USA and Alejandro Tiara from Spain presented anecdotes and a lively interchange of views giving a vivid picture of IEA's rich history. The moderator was Tom Loveless, the General Assembly Member representing the USA.

**Tom Loveless - USA:** I think we are ready to begin. Thank you all for coming this evening. As you know it's IEA's 50th anniversary, and we thought an event that we would all appreciate a great deal would be to hear from some of the titans from IEA's past. So we have assembled a distinguished panel from whom we will hear tonight. I will introduce them now from the left side of the podium, as you see them, to the right - see picture. What I will be doing, briefly, in terms of the for-



Celebrating IEA 50th Anniversary: Round Table of the IEA Honorary Members.  
Tjeerd Plomp, Neville Postlethwaite, David Robitaille, Christiane  
Brusselmans-Dehairs, Judith Torney-Purta, Alejandro Tiana.

mat, is asking them questions. They have seen most of the questions in advance (laughter). We're going to try to have a fairly free flowing discussion but not that free flowing. I have asked all of the panelists to be succinct in their answers, and so if it begins to appear that a novel such as *War and Peace* is being given, then I will cut them off. So I apologize in advance if I appear rude, but they all know that I can be that way (laughter). So, from the left to the right, Tjeerd Plomp, Neville Postlethwait, David Robitaille, Christiane Brusselman, Judith Torney-Purta and Alejandro Tiana. And we will begin—and I've asked them each to begin—with a question that they will spend two or three minutes on...and that's just their earliest memories of IEA and how they got involved. And we start with Tjeerd Plomp.

### **Earliest Memories of IEA and How They Got Involved**

**Tjeerd Plomp - Netherlands:** Thank you Tom. It is very special to have this opportunity, to share some of my impressions with you. My

first memory is in 1978, the Tokyo GA meeting . The IEA General Assembly delegate from the Netherland at that time, Egbert Warries, told me that this would be mainly a meeting for the Second International Mathematic Study. He said you are a mathematician, so why shouldn't you go to that place. Well if I am offered a trip to Tokyo I will kindly accept. I flew from Brussels and the guy sitting next to me was the Italian General Assembly member, Aldo Visalberghi, and we traveled together. Another memory to share with you is that the person who took my hand for my first trip in Tokyo was John Keeves - he wanted to go down town Tokyo and said why don't you join me, and I was very glad that I had a guide. So my first memory of IEA was of a NRC meeting and a General Assembly meeting and some very nice people that became friends for a lifetime.

**Tom Loveless - USA:** Neville is going to skip that question, because we heard his speech yesterday to the Plenary, so we go right to David...

**David Robitaille - Canada:** One of my first memories is, as Neville told us yesterday, going to a meeting and there was Ben Bloom in one corner. Having people of that stature and caliber at a meeting made a difference. Of course it also helped that my first meeting was in Paris, we were focusing on the second math study. Jack Schwille was there, Bob Garden, Skip Kifer, I think Richard Wolf. Some of my recollections of that meeting focused on the fact that there was this very mysterious way of arriving at decisions. In Canada we attend formal meetings and follow Robert's Rules of Order, at least to some degree. IEA decisions seemed to emerge at various times after local meetings that were being held. That was kind of interesting. After this General Assembly meeting a group of us was working on the math study under the leadership of Zoltan Bathory (from Hungary). There was then a very real Iron Curtain, and this meeting on items seemed to get along very well in spite of it. We had some items submitted formally and some quite informally. But we got them into good form.

**Christiane Brusselman - Dehairs - Belgium:** I shouldn't have been

the one to go to the Paris Meeting, but my boss should have gone there. In a certain way I was lucky, because he didn't like travelling, so he asked me to go and of course I said yes. I remember I was alone, I didn't know any one. My first memory, I was sitting in the breakfast room and I remember a man coming and sitting next to me. It was strange for me because I remember my mother telling me to look out for strange men and then I ask politely to him "Are you going to the General Assembly meeting?" and he said "Yes I am Bloom, Ben Bloom!" and I thought "My God, Ben Bloom is here! Ben Bloom is sitting next to me, eating eggs!" you must know that at that time Ben Bloom was, in fact he's still, very famous and all of the students had to read his books about taxonomy, his books about mastery learning. As Neville said, he was a demigod. And then the meeting started, and I remember I was very impressed about the chairmanship of Neville. From that meeting I remember one man, Neville! So these were my first impressions.

**Judith Torney- Purta- USA:** My first contact was in 1967. So that's forty-one years ago. My boss couldn't go to a meeting either, so I was invited to this idyllic lake, in the middle of New York State where they did such things such as "chart the appearance of the first robin each year on a page that went back to 1900" I was a brand new Ph.D. from the University of Chicago not from the comparative education department but from the department of human development, so I knew Arnold Anderson but not well. I felt really good when, after my presentation, he said that I was a good representative of the University of Chicago. I noticed he didn't say that until after my presentation. I had just finished co-writing a book called "The development of political attitudes in children". And I am sure that when they searched through whatever data bases were available in those days, I was I was one of the few people who had actually done empirical research on what those in IEA were calling civic education. I do remember also something about Ben Bloom and it is related to the fact that Neville talked last night about my personal quality of being determined. In one of the parties at this Lake Mohonk meeting I heard Ben Bloom say to some-

body else across the room “Oh IEA will never be able to succeed in a study of civic education!” And you can guess what my reaction to that was: “I will show him that it can be done! I will do it!” Of course I am still doing it all these years later.

**Alejandro Tiara – Spain:** Well, in comparison with all these distinguished colleagues in this table, I am a newcomer. Also I come from a different perspective. Even if I am a university professor, I didn’t start my connections with IEA as a researcher but as a member of a Ministry of Education in what it probably was a new phase in the development of IEA. I started to work for the Ministry of Education of Spain in October 1989, just after the General Assembly in Korea. My boss then and now, Alvaro Marchesi, went to Seoul and decided that we had to develop evaluation in Spain. That was my main task in the Ministry, so I got in contact with IEA since then. The first General Assembly I attended was in Beijing in 1999. It was an interesting and special moment, because there were some opinions against meeting in China. It was after the Tiananmen events and some people did not want to keep the GA in Beijing. Finally the GA was held there. But it was a small meeting in terms of people attending it. We were hosted at a researchers’ residence of the University of Beijing and really a bit isolated, like in the middle of a big canvas. But it was very interesting for me because I could had conversations with some students and other people from the University. I was also a bit puzzled, as David said, about the GA decision-making processes and so on, with its rules and procedures. When I was Chair, I did always try to explain to the newcomers how it works, avoiding the feeling of a magic experience. So this was my first contact with IEA which still continues as you can see.

### **The Biggest Obstacles to Getting the Studies Up and Running**

**Tom Loveless – USA:** We all know that in the early days there was skepticism about an international assessment, whether it could even

be done. What were some of the biggest obstacles to getting the tasks up and running? Christiane?

**Christiane Brusselman-Dehairs - Belgium:** According to me, of course it is my perspective, and maybe I will make exception of the northern countries, but I think it was the way Europe was thinking at that moment. My idea is that the IEA wanted to launch a product on the market and the market was not yet ready to accept that product. We saw that at several levels of the society. I remember going to the Department of Education and trying to get money and the general inspector, who was in charge of all the curricula, told me "Are you really sure that Belgium is going to take part in the international studies? What would be the impact? How will the minister react?" They really didn't like it. At the level of school it was nearly the same reaction. You have to know that our schools have a large level of autonomy. The teacher is the boss in the school. He gives his lessons, closes the door and that's it. And of course these teachers thought "Mmm... if international tests come in, what would happen to our autonomy." So at that level it was also a problem. Then of course at the university level, I had freely the impression, at that time, that there were enormous gaps, in the technical knowledge of European universities. At that moment, we were thinking about the philosophy of education. Elsewhere they worked with empirical data. Now of course the gap has become less important. But all of these things indeed had an impact. First of all it was hard to find the money for international studies. We got it, but it was not enough. And secondly, it was really hard to get schools to participate in these studies.

**Tom Loveless - USA:** Neville? The major obstacles?

**Neville Postlethwaite - England:** Well at the very beginning, money and also persuading some countries that this was not an American trick with American tests, especially because they were multiple-choice. Several countries didn't like that. And trying to get good sampling frames was very difficult in some countries, at that time. And

finally it was to persuade some countries that that probability samples were better than samples of convenience where they would pick the schools they wanted.

**David Robitaille - Canada:** Of course funding, but as I think about this question, I try to think back to the climate of the time of the late seventies, early eighties. There was a lot of interest in assessment. These were the early days of national assessment or developing assessment programs. So that part of it was there, but in the universities in Canada there was a decreasing interest in empirical research in the faculties of education. And this was true in the mathematics education community as far as the second math study was concerned. A Dutch scholar had really stirred things up with his reactions to the first math study, accusing the Belgians. "They did so well because the items were put together in Belgium." There was a lot of resistance in the mathematic community around the world, to the idea of an international study, especially concern about things like translation and whether you could, in fact, give the kids the same tests in different languages. So all of that was going on at the same time. They were all serious obstacles, but I am sure money was the biggest one. I certainly recall the late Roy Philipps. He went around the world; I don't know how many places he must have visited to try to raise funds for the second math study. And that was a very tough task.

### **IEA's Accomplishments over the Last Fifty Years**

**Tom Loveless - USA:** Of course the obstacles were overcome and at least IEA got started and the assessments were given. As you look back over the last fifty years and reflect on IEAs accomplishments what stands out to you. What do you think IEA and its assessments have done in term of contribution?

**Tjeerd Plomp - Netherlands:** Let me first say, I started as chair in 1989, with the task to turn the IEA into a professional organization that had the professionalism and the facilities to conquer itself and the



world. I would like to say, the major accomplishment before 1990s was that studies were actually completed and that the methodology was developed. I think we need to have great appreciation for the early generation of IEA. Those studies, in my perception, put the relevance of indicators of education achievement and other indicators on the global map. And that meant that around 1989-90, the world was ready for a professional organization. I think the fee at that time was below 200 dollars and if you compare it to figures that we have now, you see what an accomplishment it is, to have IEA turned into a professional organization.

On the cycle of studies: from 'nothing' the IEA developed the understanding that you need a cycle of studies to learn from previous studies, to have benchmarks, etc.

Another accomplishment, I want to mention is the structure. I am very proud that while I was chair the Data Processing Center was established. When the German IEA group at the University of Hamburg said 'we can no longer support this unit', the IEA had to decide, "Are we going to let it go or do we have to courage without a business plan to say 'This is a valuable group, this is important knowledge. Let us take the risk!'" I remember the General Assembly in Yogyakarta in 1994, when they took this decision. Heiko Sibberns (from the DPC) lent me his tie - first time that I forgot it. I was really nervous coming in for these decisions.

A last comment I want to make is about IEA's high level of capacity building around the world. During my chairmanship I travelled a lot and I am really impressed by the way IEA developed strategies and tactics to build capacity. Later on this responsibility for capacity building was taken by other organizations as well.

So IEA's contribution can be summarized as the studies completed, the methodology developed, putting the relevance of indicators on the global map, building capacity and developing from a bunch of motivated professors into a professional organization with experience.

**Judith Torney-Purta - USA:** I want to say something briefly about

IEA's structure. Last year I was asked to do a survey of twenty-six international projects in the United States. I was asked by our National Academy of Sciences to study collaborative processes in international research in the behavioral and social sciences. We had a quite interesting range of projects; the psychologists who had recently run IEA projects -- myself and Jim Stigler ... completed surveys. It became clear to me as I compared our responses to some of the other projects, how much better the collaboration processes was working in the IEAs projects than it was in many other kinds of collaborations. I attribute that in that, in part to the TEG (Technical Advisory Group). The other international collaborative projects surveyed were sort of inventing their projects' decisions about technical issues as they went along. These other projects often reverted to what the American project leaders wanted to do because there was no forum for technical discussion with international experts. In IEA the international TEG can give you ideas and resolve the conflicts or disagreements that might arise. The other thing is capacity building especially of young scholars or the next generation in these areas. The IEA projects, not only the ones associated with the universities, have provided enormous training to young scholars who then are going to be the General Assembly's delegates and researchers of the future.

**David Robitaille- Canada:** Well agreeing with the things my colleagues have said so far, something that is very important to me and that I appreciate in IEA's studies is the emphasis on schools and on classrooms, on curriculum, on teaching, on trying to look at students' achievements in a way that can be translated into recommended practices for teachers and curricular developers. I am glad that the IEA maintained these stands over the years, and hope it will continue.

### **How IEA Results Have Influenced Policy Over The Years**

**Tom Loveless- USA:** Of course, one of the things that often comes up in our small group discussions and in our discussions in plenary ses-

sions is the challenge of influencing policy. How has IEA influenced policy? How have IEA results influenced policy over the years? Judith?

**Judith Torney-Purta - USA:** I won't make a catalogue of those, because probably I am not the best person to do that. But I want to tell three brief stories. I remember in Washington a meeting called at the National Academy, after the second math study. Harold Stevenson was a psychologist, who had done several international studies, and he spoke at that meeting about this topic. He said "Organizations too frequently collect the data and it is something like saturating a dish rag with data instead of water. And they figure that if they squeeze out the dish rag, the policy implications of the data will drip out and fall on those who need to know." I think that maybe we haven't devoted enough attention to that. Some limitations on IEA's studies are that we still use them as a way of saying "Oh! How terrible! Our country is doing very badly! What are we going to do about our lower positions?" We need to pay more attention to the processes and to the organizations within a country that are actively interested in policy. In the civic education study one of the places where that appeared to be happening was Chile. And so we did a special report on the Latin American countries using our CIVED data. Chile was planning for a curricular reform project in civic education, having had the same curricula they had when the dictatorship was in place. It was exciting to see their reactions when they looked at some of the data, such as the fact that the television's news was trusted more than elected officials. But more importantly the analysis pointed to the kind of steps that you might explore in terms of teacher training in order to have more open classroom discussions for example. One last thing, when I started working on the first civic education study in the early 1970s the woman's movement was a kind of nascent idea in people's mind, not really prominent. So I thought it would be interesting to have a woman's rights attitude scale in our 1971 survey. We still use this scale. I should say to Neville that I believe you referred to this as "Judith's

scale." As women's rights came forth as an issue, it was a very important thing to know about how young people were reacting.

**Neville Postlethwaite - England:** Well I know too well that I am always being asked "What's the policy impact?" or "Prove that you have a policy on that!" They want something immediate. There are many books written on the relationships between research and policy making. If you remember Carol's Weiss's book, she had many different ways of dealing with it. There is not a direct relationship between research and some policy action. Rather the results from a particular research trickle down and influence the debate. The next time that you do a study, there are more people that are aware about what they could get out of it, and that is more powerful than the immediate effects. There are of course many immediate effects also. There has been policy impact in Australian in terms of social class, curricular preference, and time of school leaving between Australia's states. In terms of practical issues there has been an emphasis on processes such as opportunity to learn time.

**Tom Loveless - USA:** And we didn't always have evidence on each of those things even though they were controversial.

**Tjeerd Plomp - Netherlands:** Yes, I want to build on what Neville said, with references to Carol Weiss. There is no doubt about the relationship between the studies and the policy. But the fact that studies have been conducted and reported may in itself cause an effect. For example, there is (I think it was Neville who mentioned it) the 'mirror' function of our studies - i.e. each country may look in the mirror and ask "Am I happy with how I look?" - and the 'benchmark' function. And we know examples of countries like Australia and similar stories from Hungary after the First and the Second Mathematics Studies that showed that these studies had made an impact. I agree with the conclusion of Neville, that this is often not a direct impact, although at the international research conference in Taiwan (2007) we had very nice examples presented to us about direct impact of IEA studies.

**Alejandro Tiana - Spain:** Yes, I am of the same opinion. This afternoon we have seen several examples about PIRLS, about how PIRLS has affected educational policy in some countries. So I think the answer is “yes”. It has affected, but in this indirect and soft way probably. I have had the experience from both sides... educational research and educational policy making. And for me this is something very interesting and very difficult at the same time. It is not easy at all and I know that among you there are people more or less in the same position as I have been. I remember last year, when I was in the Ministry of Education and we presented in Spain a report on the situation of our education system. We worked a lot for making sense of what was going to be presented to a journalists’ audience and some other people. I think it was a balanced appraisal of the situation of Spanish education. At the end of the meeting a journalist came to me, after that, and said: “What you say makes sense and is very logic, but there is no news inside it. Of course, good news was no news for him and for some of his colleagues. And this is something not easy to deal with. And an international organization like IEA has to deal always with that phenomenon. I did appreciate the efforts done by the organization to translate the results of the studies into words which make sense not only to researchers but also for policy makers and other audiences. Because it can help persons in the position like mine in that moment, which I said is not always easy. But I am sure that if you look back...definitely IEA’s studies have influenced the way in which policy makers think about education and I consider it very valuable.

David Robitaille- Canada: Well, perhaps a bit of cold water on this. One of my disappointments in this area is that we haven’t done enough to reach teachers in the studies, either communicating the results to them or offering them information in some easily accessible useful format that they could get some benefits from.

**Neville Postlethwaite - England:** I don’t know to what extent researchers should also get into preparing materials or persuading

countries to take action. We will be pushed. "Well you have got all the results, why don't you now train the ministries to implement them?" Well we said "But that's not our skill. We are researchers; you are asking us to do something which we were never meant to be asked for." Maybe some other body could take this on. I was recently with a minister of education who used to be a psychologist doing research on decisions. So we said to him "Now that you are a minister what is the relationship between research and policy making?" He said "None! Look, my work it's to keep everybody happy and people have different ideologies on this, about what is good for education. If some research happens to support that, they will use it but there will be others against it. If everybody agrees to use the research it will be fine. But then do we have the money to implement it?" It is a different way of looking at life.

**Christiane Brusselman - Dehairs - Belgium:** Recently, we had a new beautiful example of use of the data. The minister looked at the data, he asked us to come to Brussels, he discussed it with us. He didn't only use one database, he used three databases and he also used his own statistics. What was interesting was the international benchmarking, especially pupils who had low achievement. When we look today at Belgium, we always look at the surrounding countries rather to watch all of Europe. We saw that the percentage of low achievement was the same as in the surrounding countries, but we saw that we achieved lower than in other countries. So the minister focused on the population and he tried to know more about the characteristics of the population. So what did he learn? It was an ethnic population, did not have a number of home resources, did not know the language. At home they spoke their home language that was different from the school language. So knowing all that, you must know he's a socialistic minister and also is minister of work. So the percentage of the population that is unemployed is of concern. And he took a number of actions. In order to enable the children to learn the language of school more easily and more early, he communicated to the parents that the kids must

go to school the first year of the kindergarten. One third and maybe more is not going for the first year in the kindergarten, so they lose a whole year of learning the language. Then he tried to take a number of regulations to persuade parents that primary education will cost nothing, also to protect that group. And I think using IEA data and all other sources of empirical research in this complex understanding is very valuable.

**Tjeerd Plomp - Netherlands:** Related to policy influence, I remember, in the early days of IEA there was a discussion about undesirable influence of policy. And if I am correct the IEA did not like rankings, neither discussions about rankings. If you look at the publications from the Second International Mathematic Study there are no rankings presented, because as a research association you don't want to send the wrong message. But what happened was that smart people started to construct rankings in a way that was undesirable. In the early years of my chairmanship of the IEA, we decided that if people want ranking, let us make sure that they do it in a correct way. Instead of publishing research reports four - five years after data collection with a complex analyses, we decided that we will have quick, policy relevant reports - univariates reported in a way that you can explain to people what it means, especially if you use confidence intervals and standard errors. So there was a sort of negative influence and wrong policy effects, but this was turned into good quality reports of univariates. Yes, it is ranking but presented in a new way.

### When IEA Came of Age

**Tom Loveless - USA:** When did IEA become of age—to use an American phrase—when did it become apparent that IEA's results were acceptable and were being accepted around the world as legitimate?

**David Robitaille - Canada:** For my answer I focus around TIMSS,

which started out as the third math study and a third science study to follow soon after. But not very long after that, the idea was put forth that we had to put the two together, and that opened up major funding from the Canadian government and from the US federal agencies. It opened the gates in terms of interesting the countries. We were expecting twenty or twenty-five countries in the third math study, as we had had in the second math study. The number soon was over forty and there were new expressions of interest almost every week. The thing that really impressed me from the Canadian perspective is that when the TIMSS results were released, in the Toronto Globe and Mail the top third on the front page, what they call the flare, was about IEA's TIMSS results. This was unheard of and perhaps even more amazing, was that in that newspaper there were two editorial cartoons about Canadian performance in this international study. The results had a massive impact, I thought. There was a comprehensive article in the Economist, on TIMSS results. Well, that was an impact especially coming some time after the study. We were getting the sort of recognition of the studies that Tom's question implies.

**Tjeerd Plomp - Netherlands:** Let me built on the first remark of David. We decided to do the Third Mathematics Study at the General Assembly in Korea (*TIMS; in 1989*). I started my chair period after that General Assembly and I still remember that the first AERA conference after this was in Boston, perhaps 1990. I was invited to a meeting called by Emerson Elliot, at that time, the Commissioner of Education Statistics in the US. He told me that an international cooperative study (*IAEP, International Assessment of Educational Progress*) organized from ETS in the US met the criticism that it was perceived as an US study and its comparability was not convincing. So the US Board on International Comparative Studies in Education at the National Academy of Sciences wanted an established international organization conducting studies to give the US a valid picture of educational achievement. So in that particular meeting, I was asked, "Are you willing to broaden the scope of the TIMS to make it the TIMSS Study (*i.e.*



*the Third International Mathematics and Science Study*)? If you do that, we will support you, but I need a reply from you!" I still remember that I knew the rules of my organization and I knew that my Standing Committee should be consulted, but at that time I was forced to take a quick decision and I agreed. The study turned very well. But for me that was a sort of a milestone for IEA. It was recognized by an important member country as an organization that can do studies.

Let me give another example. When I was invited to become IEA chair, linked to the invitation was the question "Is there a possibility to find money for establishing a secretariat?" I was not the only one who received that question – it was sent to a number of people. But it was interesting to see that a number of governments like the Dutch government were willing to put up a budget to allow the IEA to establish a Secretariat and an office. I think that around late 1988 and early 1989, the reputation built by scholars in the early days was such that the policy makers recognized the importance of the work and we had a chance. And I think as an organization we can be proud that we had this opportunity and that we made it happen.

**Judith Torney- Purta - USA:** I would target the same period. I remember, Tjeerd, that you came to AERA that year for a session I had organized, and that you had to throw away your old transparencies and handwrite some new ones because the decision to go ahead with a study of math and science together was so recent. But I want to give a little bit of perspective on BICSE, on the Board of International Comparative Studies in Education, which was at the U.S. National Academy of Sciences, which was advising the Commissioner on these issues. It was established because Dorothy Gilford, who was part of the statistical group there, was concerned about response rates in the Second Science Study and wanted to get some statisticians and educational researchers from a broad range of US universities and agencies to talk about this. When the TIMSS project came along, somebody quipped that "Overnight IEA went from a cottage industry to be the equivalent of the Manhattan project." And BICSE was an attempt on

the part of the Academy and others to regularize things so that materials for these decisions didn't come from one person's memory. BICSE provided a chance to hear and see various plans that were made and how they might be improved. At the beginning of TIMSS there was a serious probability of it not succeeding. The money wasn't there, and Al Beaton was tearing his hair about items and budgets. Now we look back and we say "Wow! Great success!" but there was a time that it was on the edge and a lot of people argue that BICSE played a positive role.

**Tom Loveless - USA:** Anybody else want to... Neville?

**Neville Postlethwaite - England:** I guess I have a different perspective because I think a turning point was to identify the malleable determinants of achievements, which were more important and which less important. When a challenge came up, IEA met it. One of the strengths of IEA is that it is able to live up to demands and just do it. The third age came with multilevel analysis, and again there always was the idea to identify educational variables, which are important. And then the next age was moved mainly by economists wanting to predict economic development at later times from achievement scores in earlier times. IEA has come of age many times for me as it has met these challenges. I am not interested in the coming of age when it means recognition by the press.

### **Future Challenges for IEA**

**Tom Loveless - USA:** Well you have actually segued nicely to our last question. And then we will have a little bit of time. So those of you out here who have questions you'd like to ask, think of them now. Our last question is... Looking out in the future over the next fifty years, what do you see for IEA, what do you think IEA should be doing, what do you recommend? Or just generally what do you see as future challenges for IEA?

**Alejandro Tiara - Spain:** Well I think from the comments of my colleagues around the table, we can conclude that IEA is a history of success. I know that things have always lights and shadows. There always are things to be done, to be improved. But looking at the fifty years, I think that a number of challenges were faced and were done in a successful way. I think we have to thank all the people who have been contributing to that. I will recommend just to keep an eye on what IEA needs for its development. I think what it has done in the recent past with development for instance of the Data Processing Center in Hamburg, the Secretariat and so on. By going into more depth, more substantial issues, for me, I will recommend to continue this line of dialogue and cooperation among countries. I think this has been successful. IEA did face, in the past, the risk of being considered as a tool of one country or a small group of countries. I think that things may develop in another direction, enlarging participation, having countries from different regions of the world, which in fact are a balance of countries. This means having wealthy countries, countries not that wealthy, countries having a long history in educational research, countries coming recently to that field. And I think that this mixture is something, which is one of the powerful engines of IEA. And I will recommend keeping that, not to leave out countries especially in Africa, for example. I think this is difficult but I think more has to be done. Another area, which is much close to me, is Latin America, which in fact has started a level of representation in IEA studies. And the last recommendation is to keep this balance which is going on now between reliable research, good studies, studies done in time and at the same time responding to key issues in education. This means that the organization has not only to keep doing what it knows to do but also to explore fields that may seem less central. From this point of view I would like to end saying that for me, the studies of civic education are a very good example. At the beginning many thought that it would not be possible even to do that area. At the end having developed the knowledge and a set of instruments in this area has proven at the beginning

of the 21st century to be addressing a key issue for the development of our society and the educational system.

**Tjeerd Plomp - Netherlands:** I have come to the conviction that increasingly we should look at our studies as core studies to which national options and regional options should be linked. If you look at the list of indicators the EU is interested in or Unesco, I can imagine that organizations like these want to join IEA studies – maybe that is already the case; I am not informed about the last developments.

As a second point I want to mention is that there is a whole group of researchers in the world working on school effectiveness, on school improvement. I know the IEA pays a lot of attention to background questionnaires. I don't know whether people in school effectiveness and improvement and the IEA are working together. If that is not the case, then my recommendation is that it should be done.

Nowadays in a lot of policy discussions we talk about "Learning to learn," "Lifelong learning" and things like that. I think is important the IEA incorporates these concepts in its studies. Maybe you do it already, but from my present position I can't see it, and that's why I want to mention it.

There are new topics. We tried in the 1990's a study of foreign languages but we couldn't find the funding. But when I'm looking around in Europe, I see how important languages are, and the IEA did it in the past. It might be an idea for a future topic.

There is an excellent IEA website, and I use it very often. But sometimes I like to have the old fashioned newsletter twice a year, just to get informed about the developments. It may be sent electronically, but as an old man I really miss to be informed regularly about developments in IEA.

And the last point, this is closest to my heart. I really enjoy the international research conferences and I think it is excellent that the IEA has taken this initiative But I want to make a plea that the IEA also goes

out to other research conferences. Some of you know that at the European Conference on Educational Research already for ten years, a group of people organises a symposium on international comparative studies and analyses of studies like TIMSS, PIRLS, PISA, and at the last conference we had about twenty papers. If I look at the attendance, there were people who are not usually participating in the network of IEA. So I can imagine the IEA makes it a strategy to assure that at the regional conferences or international conferences this sort of analyses be presented. I think is an important way to inform people of what is happening in IEA studies.

**Judith Torney- Purta - USA:** Just a couple of comments relating things which have been said. First of all about the importance of context and background measures. We found our two phased study with a case study at the beginning to be absolutely invaluable to us in anchoring what we were going to study in real experiences of schools and the real issues which are there. TIMSS did a case study after they had finished much of the analysis. Gerald LeTendre and others argue that maybe it is better to do case studies first. It's messy, it's time consuming but it also allows people who are working on these studies in the countries to have an intermediate product; they don't have to wait for the final report, but halfway through the project they produce something of value. The other comment is related to what was said about the balance between researchers and policy makers. When I was in Australia presenting the results of the civic education study, I reported that Australia was one of the few countries in which the students who said they wanted to protest were also the ones who really had skeptical attitudes that things should change. So after my speech was over, an Australian teacher came up to me, and said, "I'm so glad to hear that we are not losing the Australian edge!" Applying this to IEA, I hope that as this organization becomes well known and financially successful, that we don't lose our innovative edge.

**Neville Postlethwaite - England:** In the long paper which I distributed, I have a lot of things to say about the immediate future and some

of the technical issues. The other thing I would say like to say, which maybe is a bit controversial, is that, I have a feeling that there are too many international studies and schools are getting over tested, especially in small countries. This is beginning to be true everywhere (not just in Europe and the U.S.). I don't see immediately how to solve this problem, but I think it should be examined.

**Tom Loveless - USA:** Before we go to questions, let's thank our panel not just for this evening, but for their fifty years of contribution to IEA. Thank you very much.

**Question from the audience:** Thank you very much for this very interesting discussion. I have a simple question, "You have not mentioned students. What's the responsibility of IEA to the students and their future?"

**David Robitaille - Canada:** Well, I thought I did mention it. That's what I intended when I said, I think that the IEA needs to do a better job, than it does, to communicate with schools and teachers about the results of the studies. I don't know if Neville agrees.

**Neville Postlethwaite - England:** The whole reason for doing the multivariate analysis is to identify the variables which could be used to help all students. Now, I do think that in some cases countries need a different model. In some cases, some populations of students need to be oversampled from the start, to be able to do the analysis later. But the basic aim of the studies is always to help the students!

**Tom Loveless - USA:** Thanks once again and thank you all for taking part!



## CHAPTER 22

# **The Scientific Contributions of Torsten Husén and Neville Postlethwaite to the Development of International Comparative Research on Educational Achievement**

*Rainer Lehmann*

### **Torsten Husén and T. Neville Postlethwaite: Two Impressive Personalities in Comparative Education**

In 2009, the International Association for the Evaluation of Educational Achievement (IEA) lost two of its most renowned representatives within a period of just a few months: Torsten Husén (1916-2009), late Professor of the University of Stockholm, and T. Neville Postlethwaite (1933-2009), late Professor of the University of Hamburg. Far beyond their importance for the Association, both men contributed substantially to the development of international comparative research in education as an academic discipline.

On the occasion of the Conference of the Comparative and International Education Society in Chicago, in March 2010, a small number of Torsten's and Neville's close collaborators and friends gave their very personal accounts of the role that these truly exemplary researchers played in the academic community in general and in the IEA in particular. In two separate addresses, Judith Torney-Purta and John Schwille illuminated Torsten's and Neville's "intertwined and extraordinary careers", as Dr. Schwille phrased it; Ingrid Munck highlighted Torsten's exceptional qualities as an inspiring "doctor-father"; and I tried to highlight Neville's growing, truly seminal influence on educational thinking and policy making in Germany and other countries around the globe.

In an attempt to illustrate the significant scholarly contributions of Torsten Husén as a leading member of IEA's group of 'architects', and those of Neville Postlethwaite as the 'chief engineer' during key stages



of IEA's development, a fair amount of biographic detail, is, indeed, inevitable. As always, a combination of personal traits and fortuitous circumstance facilitated the remarkable influence these two men exerted over the Association during its prime years.

### *Torsten Husén*



Torsten Husén came from a relatively humble background – his father was a sawmill manager, his mother had been trained as a telegraphist; his maternal grandfather was a primary school teacher, his uncle headmaster of a special education school. It is all the more amazing to note how Torsten, through his remarkable personality and talent, developed this heritage straightforward and very rapidly into a wealth of 'human capital' from which the Association profited for many years.

After having followed the traditional schooling of the gymnasium in Växjö in Småland, he was accepted for university studies, where he made a significant investment in mastering three foreign languages. In fact, the only recorded exception to Torsten's habit of attacking supreme challenges via the 'direttissima' method was his omission of sustained French studies, later compensated by hefty efforts during summer vacation. In any event, Torsten managed to convert these opportunities to learn – and all subsequent opportunities throughout the years – into professional resources for educational research in general and for IEA in particular.

The steps leading to his amazing accumulation of expertise can only be mentioned in brief here.

- first university studies in Lund (1935), beginning with courses in mathematics, history and literature and then moving on to psychology – this breadth of interests is remarkable, indeed;

- involvement with the impressive Malmö Study (1938/39), concentrating on the relationship between endogenous and exogenous (i.e., social) determinants of learning achievement (Fägerlind, 1975);
- completion of his studies in developmental and differential psychology, and obtaining his licentiate in psychology with a thesis on eidetic imagery (1941);
- work on a system of psychological tests for the Swedish Armed Forces (beginning in 1942);
- completion of doctoral studies with a monumental dissertation on adolescence in Sweden (1944);
- extension of research related to the military, continuation and generalization of investigations of the interaction between 'nature and nurture', initiated with, but not limited to the Malmö Study; (A remarkable example of this type of research is the Twin Study, based on Army data collected between 1948 and 1952 and published in 1953);
- broadened scope through philosophical and historical studies, namely treatises on Fridtjuv and Anders Berg's efforts to promote a unified compulsory school system (1946, 1948, 1949).

This truly impressive corpus of psychological research, maximally relevant for educational theory and practice, led to a series of important academic appointments:

- Professor of Educational Psychology (University of Stockholm, 1953)
- Professor of Applied Educational Research (University of Stockholm, 1956)
- Professor of International Education (University of Stockholm, 1971)

Torsten Husén's renown as an incredibly productive researcher and

his academic positions were the assets which could be converted into 'social capital' in both the academic and the political arena.

As early as 1952, Torsten was called upon by the US High Commissioner in Germany as a consultant on educational policy – incidentally one of the few apparent setbacks in Torsten's career, given the pronounced, but frustrated American interests in establishing a comprehensive school system in that country which had not yet regained full sovereignty after the Second World War. It may also be noted that this initiative led to a co-operation with the German Institute for International Educational Research, the institution which then (and again since 2008) functioned as the German representative in the IEA General Assembly.

Visits to the United States – of increasing length and importance – followed, among them a crucial visit to the University of Chicago where, among other leading educators, C. Arnold Anderson represented the field of Comparative Education and where Benjamin Bloom had just begun to systematise and modernise curricula by introducing his famous "Taxonomy of Educational Objectives" (1956), used, among other things, as a tool to buttress "mastery learning" (Block, 1971). It was in this intellectual environment that the idea of implementing cross-cultural, output-oriented educational comparisons arose. And it was none other than Torsten Husén who added the idea of considering "the world of education as a natural laboratory in which different countries were experimenting with different strategies of teaching and learning"; considering it as an experiment, so to speak, that enabled us to study the effects of the various observed 'treatments' (Husén, 1967; Keeves, 2004).

With the benefit of more than half a century of experience gained from methodological trials and errors – and, of course, with the proverbial wisdom of hindsight – one may be tempted to consider the early efforts to establish such an innovative strand of research as overly optimistic and to frown upon some of the early simplifications.

Two points should be borne in mind, however:

First, it must be understood that, from a pragmatic point of view, such optimism was absolutely necessary, if there should be the slightest chance to assemble the required international expertise, to secure the funds needed, to put into place the institutional arrangements, to obtain the necessary political support as well as the compliance of many thousands of teachers and students without whom there would have been no comparative studies at all. It was this optimism – Torsten’s refusal to acknowledge seemingly insurmountable difficulties – which Neville Postlethwaite believed was, indeed, the key both to Torsten’s amazing academic career and to IEA’s success.

Second, many of the technical standards which are now considered essential not only for international comparisons, but far beyond in the area of any large-scale assessment of educational achievement, could only be developed on the basis of actual research experience – on board a moving ship, so to speak. The book outlining the formulation of “Technical Standards for IEA Studies” (Gregory & Martin, 2001) as well as the International Handbook on “Educational Research, Methodology, and Measurement” (Keeves 1988; 2nd ed. 1997) demonstrate the effect which the decision to embark on quantitative international comparisons in education has had. Incidentally, the Handbook was largely compiled from articles written for the monumental “International Encyclopedia of Education” edited by Torsten Husén and Neville Postlethwaite (1985, 1994).

A word needs to be said here about the institutional steps taken by Torsten Husén and the other ‘founding fathers’ of IEA in the late 1950s and early 1960s to organize internationally comparative research as we know it today:

- 1958 In a famous meeting at the UNESCO Institute of Education (now Institute of Lifelong Learning) in Hamburg, the decision is taken to conduct a Feasibility Study in Hamburg. (Inasmuch as this is considered the ‘birthday’ of the

Association, the General Assembly in Berlin 2008 marked its 'half-Centennial'. Therefore, Neville Postlethwaite was invited to reflect on his experience with IEA research. The fact that this turned out to be one of his last public appearances underscores the nature of this speech as the formulation of a heritage to be maintained and developed.)

- 1962 Torsten Husén is elected Chair of IEA (until 1978).
- 1964 The First International Mathematics Study (FIMS, not yet known by that name at the time), is implemented with key activities already taking place under the auspices of Torsten's Chair at the University of Stockholm and with significant contributions by Neville Postlethwaite (publication of results in 1967).
- 1971 The comprehensive IEA Six Subject Survey is conducted at Torsten's new Institute of International Education at Stockholm University. (For a summary beyond the thematic volumes, see Walker, 1976.)

### *Neville Postlethwaite*



Like Torsten, Neville attended a grammar school, and also like Torsten he spent some time in military service after secondary school (1951-1953); unlike Torsten, however, his military service was not obligatory, but Neville's family background did not allow him to take up university studies right away. To shuttle continuously in slow airplanes between Baghdad and Hong Kong for almost two years in order to save money for university is certainly convincing proof of high educational aspirations. Eventually, Durham University became Neville's Alma Mater, awarding him in 1956 a B.A. in Social Studies and in 1957 a Diploma in Education,

which served as a basis for teaching as a lecturer at St Alban's College of Further Education (1957-1961) and to work, until 1962, as a Research Officer in the Test Services group within the National Foundation for Educational Research in England and Wales (NFER, long-standing member institute in the IEA General Assembly).

1963 saw Neville's move to the UNESCO Institute in Hamburg, whence he transferred rather quickly to the University of Stockholm which offered better working conditions. Neville's key role in the actual analysis of the FIMS data is to be noted here. The product from this co-operation was the classic two-volume book entitled, *School Organization and Student Achievement* (Stockholm/New York, 1967), with Torsten Husén as the primary author and editor, but also with substantial contributions from Neville's desk. In fact, this type of work formed the basis of his Licentiate in Educational Psychology, obtained from Stockholm University in 1965, complemented by the respective doctoral degree also from Stockholm University in 1968. When the Six Subject Survey was implemented, during a period when IEA was experiencing rapid initial growth, Neville Postlethwaite was part of the group of key researchers, functioning very much like an Executive Data Manager (although this position in IEA was created - temporarily - only much later).

Based on the success achieved during this period, it is easy to see that Neville possessed a stupendous energy which he devoted almost entirely to the IEA. The fact that he not only succeeded Torsten Husén in the Chair of IEA (1978 - 1986), but dealt with the ongoing - even existential - challenges facing the Association with complete success bears testimony to the truly remarkable efficiency and effectiveness of his work.

Neville's scholarly merits were fully rewarded academically in 1976 by an appointment as Full Professor in Comparative Education at the University of Hamburg.

Like Torsten, Neville succeeded in transforming his 'human capital'

into 'social capital' within the academic community, namely by establishing a truly amazing network of friends and professional partners in all fields of education. This helped tremendously in keeping pace with the rapid technical advances in educational research. In this context, it must be remembered that on several occasions, Neville took the initiative to save major IEA studies from failing:

- In 1987, he organized the completion of the Second International Science Study (SISS) when there seemed to be neither financial nor human resources at hand to achieve this. One key measure in the study's successful completion, by the way, was to assemble a team of students with advanced statistical and computer skills. These same students later formed the nucleus of what is now the IEA Data Processing Center and the joint IEA/ETS Research Institute.
- From 1988 - 1994, Neville served as the International Coordinator for the IEA Study of Reading Literacy.

When he was no longer directly involved with IEA studies and after he had retired to southern France, Neville resumed his interest in and engagement with developing his technical skills in new data analysis techniques, often in the context of evaluative studies in less privileged countries, under the auspices of UNESCO/IIEP.

Both scholars, Torsten Husén and Neville Postlethwaite, maintained extraordinary levels of productivity throughout their lives. On occasion, Neville characterized Torsten's way of life with the Latin adage, "nulla dies sine linea" ("[let] not [pass] any day without [writing] some lines"; 2001), a principle that is equally applicable to his own working habits.

### **The Concept of Productivity-Oriented International Comparisons**

As John Schulle recalled in his memories of meeting C. Arnold

Anderson, another of IEA's 'Founding Fathers', in Chicago in 1963, some comparativists at the time felt an urgent need for comparable 'dependent variables', as Anderson expressed it. Today we would, perhaps, prefer to speak of 'output indicators' or 'criteria of success in education'. Anderson knew, of course, that such evidence was about to be produced – by the FIMS, to be specific. The idea seemed straightforward and convincing:

- measure the achievement distribution in a particular domain, in a sample representing the entire target population of a given country (e. g., the students in the school grade preceding the end of compulsory schooling);
- compute one or more meaningful indicators of the aggregate attainment of the selected output measure(s);
- rank order the populations/countries investigated accordingly;
- interpret the figures obtained as sufficient statistics for judgements as to the quality of the investigated systems.

It is easy to see that the overarching idea in this line of argument was some basic notion of educational productivity. It would be a gross simplification to say that the early IEA studies had no theoretical underpinning, that they were nothing more than a reduction of the concept of educational quality to simple mean comparisons. Yet, in fact, the researchers of even the very first projects, such as FIMS, for example, realized the need to explain this .

One aspect of educational productivity that the early studies considered can be seen in the attempt to construct and interpret 'yield curves', an aesthetically appealing precursor to the current technique of comparing percentile bands. Another aspect which clearly required consideration was that of population coverage. In a famous critique of the German reluctance to acknowledge irrefutable shortcomings of the German system of education, Neville demonstrated, in 1967, that fallacies were incurred by not accounting for such characteristics in the achieved samples (Husén 1967).



Nevertheless, it seemed that the techniques for differentiating an observed achievement distribution according to major stratifying variables such as school track were still inadequate – they were insufficient to provide a deeper understanding of the quality of education systems and their underlying mechanisms.

The political context which motivated many educational policy makers of the time, including significant funders such as Sweden's *Riksbanken*, to authorize and/or fund these early IEA studies was not in the least defined by the contemporary controversies over comprehensive schooling. There can be no doubt that Torsten in particular, also being a member of Sweden's governmental School Commission, had strong interest in evidence which would support – or at least not damage – the move for comprehensive schooling, which was seen as an essential component of modernising and democratising the country. Arguably, however, the actual findings had only a limited influence on the relative legislation. At the height of the battle in the mid-1960s, the British Minister of Education (Anthony Crosland), who had consulted with Torsten, was reported to have argued that although supporting evidence would be very welcome, even in its absence, comprehensive schooling would be a political necessity anyway. It is clear, then, that educational achievement as the hallmark of productivity was not universally considered the ultimate criterion of successful educational efforts

To merely consider the relationship between the structure of the educational system and its aggregate output can be seen as the final stage in a concept of Comparative Education restricted to juxtaposing whole systems as culturally determined Gestalten. Beginning with the analyses which took aggregate input measures into account (such as Neville's approach with social characteristics and the system's retentivity as control measures), a line of research emerged which developed this concept more systematically. It seemed both possible and required now to use IEA-type research to search for reasons – i.e., for quantifiable influence factors – that could explain the achievement dis-

tribution in the systems to be compared.

Such comparisons were perfected under the label, "Educational Productivity", introduced by Herbert Walberg and his school of thought (see Walberg's 1984 seminal article). There was, indeed, quite substantial cooperation between Walberg's group and IEA researchers; so much so, in fact, that some influential authors accused IEA of being overly, if not exclusively, committed to the productivity paradigm.

### **The Search for Suitable 'Manipulable Variables': the Quest for Explanations of Systemic Success and Failure in Education**

Educational research activities, including IEA studies, have often produced evidence which is highly explanatory in statistical terms (as indicated by a high  $R^2$ ), yet rather useless for practical purposes, because the respective predictors of achievement are not subject to intervention. In fact, the productivity functions obtained often contain such 'non-manipulable' terms. That is why Neville Postlethwaite always included the quest for manipulable predictors of educational success as one of the primary aims of educational research.

The borderline between manipulable and non-manipulable variables is not always easy to define, however. Let me use the relationship of gender and reading achievement as an example.

- (1) There seems to exist a near-universal superiority of female students over males in the domain of reading comprehension (for primary schools, see Mullis et al., 2007, p.48).
- (2) Gender in itself is generally a non-manipulable variable.
- (3) It seems to follow that little can be done to change the males' deplorable situation.

Obviously, this conclusion is fallacious, if the real cause of the males' inferior performance is found in some covariate which happens to

interact with gender, such as quantitative differences in reading habits and practices, qualitative differences, e.g., gender-specific predilections for certain text genres and the like. It is clear that it is highly desirable to search intensively for such interactive effects, as these can open a way to successful intervention.

In fact, much of the current literature in educational research can be understood as a quest for strategies to transform seemingly non-manipulable influence factors such as socio-economic background, educationally relevant resources in the home or parental support into factors open to intervention.

As I have already implied, such arguments may become particularly heated when the large data sets typically produced by IEA or similar studies are used to identify relevant structural characteristics. It should be noted that this tendency has been apparent almost from the beginning and could even be considered part of the legacy of the intellectual pioneers in the field, namely IEA's leaders in that critical phase of the development of the Association. Given the enormous importance of such issues both for the quality of education and for the equity of access to it, it seems well justified to address these issues in as many different ways as possible and to discuss the findings and conclusions in academic settings such as this one.

It is a lasting contribution of both Torsten and Neville to have constantly addressed—in their teaching and training, their writings and public appearances—the relationships between possibly invariant cognitive resources and variable educational outcomes, between educational opportunity and educational achievement/attainment. Thus, it is not a coincidence that the great American sociologist James Coleman was a close friend of Neville's, and who, despite his more sceptical views on education, had a great deal of theoretical depth to add to the primarily psychological frames of reference that Torsten and Neville used.

It would seem that IEA would gain significantly if the Association

could intensify its efforts to secure the interest of outstanding scholars from adjacent fields such as Economics of Education – theory-guided researchers whose analyses could in turn profit greatly from the ‘evidence added’ by IEA studies over and above that obtained from competing programs.

### **Conceptual Foundations and Technical Advances**

Many of the technical innovations which have facilitated a lasting interest in the conduct of IEA studies are related to the activities of outstanding doctoral students who have been advised by Torsten and Neville, and who have become the new generation of great contributors to progress in education. The following is a very small selection of key words referring to advances in the analysis of IEA data which were contributed by their graduate and doctoral students:

- (1) *Confirmatory factor analysis (LISREL)*: Ingrid Munck – application to IEA Six Subject Survey (1979)
- (2) *Soft latent trait models (PLS)*: Norbert Sellin – application to Classroom Environment Study (1991)
- (3) *Multilevel analysis (HLM)*: Petra Lietz – application to the Reading Comprehension Study (part of the IEA Six Subject Survey 1970/71) and the IEA International Reading Literacy Study of 1990/91 (1996)

There are, of course, many others who could have been named here – Torsten Husén’s and Neville Postlethwaite’s direct or indirect ‘academic descendants’ who have used IEA and other data to refine the methodology of this line of research. The exponential growth of such ‘pedigree’ corresponds to the increase in knowledge and expertise which is needed to improve education and its prerequisites worldwide.

## References

- Block, J. H. (1971). *Mastery Learning: Theory and Practice*. New York: Holt, Rinehart, Winston.
- Bloom B. S. (1956). *Taxonomy of Educational Objectives, Handbook I: The Cognitive Domain*. New York: David McKay Co Inc.
- Fägerlind, I. (1975). *Formal Education and Adult Earnings: A Longitudinal Study on the Economic Benefits of Education*. Stockholm: Almqvist & Wiksell International.
- Gregory, K. D., & Martin, M. O. (2001). *Technical Standards for IEA Studies*. Delft: IEA.
- Husén, T. (1941). Studier rörande de Eidetiska Fenomenon [Studies concerning Eidetic Phenomena]. Series: *Lunds Universitets Aarsskrifter*, N.F. Avd 1, Vol. 41(7).
- Husén, T. (1944). Svensk ungdom [Adolescence]. Psykologiska undersökningar av ynglingar i aldern 17-20 ar. Stockholm: Gebers.
- Husén T. (1946). Fridtjuf Berg, folkskollärarkåren och stavningsreformerna. *Pedagogiska skrifter*, 192.
- Husén, T. (1948). Fridtjuf Berg och enhetsskolan. *Pedagogiska skrifter*, 199.
- Husén, T. (1949). Anders Berg under folkskolans pionjärår. *Pedagogiska skrifter*, 205.
- Husén, T. (1953). *Tvillingstudier* [Studies on the Psychology of Twins]. Stockholm: Almqvist & Wiksell.
- Husén, T. (Ed.). (1967). *International Study of Achievement in Mathematics. A Comparison of Twelve Countries* (Vols. 1-2). Stockholm / New York: Almqvist & Wiksell / John Wiley & Sons.
- Husén, T., & Postlethwaite T. N. (Eds.) (1985). *The International Encyclopedia of Education*. 10 vols.; 2nd ed. 12 vols. 1994. Oxford: Pergamon.
- Keeves, J. P. (Ed.). (1997). *Educational Research, Methodology and Measurement: An International Handbook* (2nd ed.). Oxford: Pergamon.

- Keeves, J. P. (2004). Monitoring the Learning and Teaching of Science in a Changing World. *International Education Journal*, 5, 275-293.
- Lietz, P. (1996). *Changes in reading comprehension across cultures and over time*. Münster / New York: Waxmann.
- Munck, I. M. E. (1979). *Model Building in Comparative Education. Applications of the LISREL Method to Cross-National Survey Data*. Stockholm / New York: Almqvist & Wiksell International.
- Mullis, I. V. S., Martin, M. O., Foy, P., & Kennedy, A. M. (2007). *PIRLS 2006 International Report. IEA's Progress in International Reading Literacy Study in Primary School in 40 Countries*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center.
- Postlethwaite, T. N. (2001). "Torsten Husén" In J. A. Palmer (Ed.), *Fifty Modern Thinkers on Education: From Piaget to the Present*, (pp. 96-102). London and New York: Routledge.
- Sellin, N. (1991). *Statistical Model Building in Research on Teaching: The Case of a Study in Eight Countries*. Unpublished doctoral thesis. Hamburg: Faculty of Education.
- Walberg, H. J. (1984). Improving the productivity of America's schools. *Educational Leadership*, 41 (8), 19-27.
- Walker, D. A. (1976). *The IEA Six Subject Survey: An Empirical Study in Twenty-one Countries*. Stockholm / New York: Almqvist & Wiksell / John Wiley & Son.



## CHAPTER 23

### **The Evolution of the IEA: A Memoir<sup>1</sup>**

*Alan C. Purves†*

The material for this paper comes from three major sources: a series of letters from Dr. David A. Walker written to me during May-October 1982; a lecture, "IEA: Past, Present, and Future," given by Torsten Husen at the Department of Education, University of Chicago (March 3, 1984), in honor of Benjamin S. Bloom; and a series of papers (alas, no longer available) prepared for a seminar entitled "Sociocultural Factors in the Management of ICRD Projects: The IEA School Subjects Survey Project," sponsored by the East-West Culture Learning Institute, East-West Center, Honolulu (January 12-16, 1981). From these documents and from the minutes and my memories of various meetings both formal and informal, I have sought to weave this tapestry and to interpret the pictures. I am grateful to all these sources and to many other people who have offered information about the life of the IEA, in particular Torsten Husen and T. Neville Postlethwaite who encouraged and at times goaded me into this writing.

To write a history of the IEA is to prepare a chronicle of an organization; so I shall try to avoid writing a history. Instead, I shall essay a memoir, something more akin to a biography, because, as so many of the sources I have consulted state or imply, the IEA is less an organization than a group of people, and it is as a group of people that it has survived and enjoyed success.

---

1. From "The evolution of the IEA: A memoir," by A.C. Purves, 1987, *Comparative Education Review*, 31, p. 10-28. Copyright 1987 by the Comparative and International Education Society. Reprinted with permission.



## **First Phase: The Republic of Founding Colleagues**

The IEA appears to have been so named first in Hamburg, West Germany, at the Unesco Institute for Education on April 17, 1961, as a footnote to the agenda of the meeting of the Evaluation of Intellectual Processes Group; "An opportunity will be taken during the week to hold a brief discussion devoted to the International Project on the Evaluation of Educational Attainment (IEA)." Such an undramatic baptism followed a lengthy gestation. An informal group of educational researchers had met at the Unesco Institute since 1955, drawn together by W. D. Wall of England to discuss various common concerns. Torsten Husen suggests that the gestation came about because of Sputnik, which made U.S. educators aware of the fact that others might do things better. Simultaneously in Europe there was an increasing interest in comprehensive schools and the provision of secondary education to a larger proportion of the age-group. The desire for change prompted a looking beyond one's borders. Coupled with this desire was an increasing interest in educational measurement on both sides of the Atlantic and, with it, an interest in quantitative and explanatory data rather than the narrative description then prevalent in comparative education.

From this ambience there emerged a proposal, brought to the Unesco Institute of Education by Donald Bigelow of the U.S. Office of Education and written coincidentally by A. W. Foshay of Teachers College, Columbia University, and Benjamin S. Bloom of the University of Chicago, for "an international study of intellectual functioning." Foshay's (1962) rationale for such a study follows:

What has not heretofore been attempted, even on a limited basis, is a comparison that would take the school population near a terminal point, and involve many countries from the same general world culture... Such an effort, however, would have advantages: the results could be examined with one's mind on the fact that they arose

from many apparently different conceptions of the nature and meaning of education; since the students were near the end of their formal education one might take the test responses as representing the outcome of the educational system as a whole, rather than catching a student in mid-career, before the curriculum had been completed (p. 3).

The purpose of the study was twofold: to make inferences about intellectual functioning from multiple-choice items and to test the feasibility of a large-scale international study.

The Board of Governors of the Unesco Institute agreed to support the international costs of the study and to convene an initial meeting held in Hamburg, June 1-5, 1959, which may be considered the actual birth of the IEA. The representatives included Fernand Hotyat (Belgium), Gaston Mialaret (France), Walter Schultze (Federal Republic of Germany), Moshe Smilansky (Israel), G. D'Arcais (Italy), Jan Konopnicki (Poland), Torsten Husen (Sweden), Douglas Pidgeon (England), and Benjamin Bloom, A. W. Foshay, Donald Super, and Robert Thorndike (United States).

Immediately following this meeting, this group met again at Eltham Palace, London, for a meeting of representatives of European Centers of Educational Research, where they were joined by W. D. Wall and A. Yates from England and David Walker and Stanley Nisbet from Scotland. Wall, then director of the National Foundation for Educational Research, chaired the meeting and provided most of the translation for the French-speaking participants. The Eltham Palace meeting, like that at Hamburg, was spent discussing issues that have since become only too familiar to the IEA: sampling, test construction, questionnaire items, timetables, and costs. Far from palatial, Eltham Palace provided dormitory rooms that the representatives shared and no soap or towels. Despite, or possibly because of this, the participants were friendly and even took time off to play croquet on the lawn and one evening to have a party with a round of song, an event that has

been repeated at subsequent meetings of the IEA.

The full results of the pilot study are given by Foshay( 1962) and others, but the main result can be summed up in the sentence, "Cross-national comparisons of educational performance can be made and can give comparable results" (p.62). Such a finding was startling at the time, but more important, perhaps, was the clear sense that a group of researchers from different cultures considering different educational systems could agree on a common approach to testing and evaluation. Foshay's original aim of studying intellectual functioning seems to have been replaced by a much more sharply defined curricular base to the test items. David Walker contributed the phrase "opportunity to learn," which came to be one of the important items for study in subsequent IEA projects, even though Walker's analysis of the pilot study data found that "native ability" accounted for more of the explained variance in successful completion of an item than did the teachers' emphasis in class.

Carrying out such a project, of course, involved meetings, and these were held in Hamburg during 1960 and 1961, with most of the same people as were at the Eltham Palace meeting. In late 1960, there was some discussion of doing a larger-scale study of achievement, and mathematics was one of the subjects proposed. Husen states that a "council for the international evaluation of educational achievement" was formed in October 1960, but Walker writes that he can find no earlier mention than 1961, with a slightly different title (as is true of so many mythical characters, the birth and parentage of IEA is cloaked in mystery). It is clear, however, that the meeting in June 1961 was a watershed.

Donald Super started off as chair of the meeting because both Wall and Husen were late in arriving. Others present were Bloom, Foshay, Thorndike, Mialaret, S. Roller (Switzerland), Schultze, Martti Takala (Finland), Lars-Masnus Bjorqvist (Sweden), and Smilansky. From the Unesco Institute were Saul B. Robinsohn (the director), Willi Koelle, and David Cobb (who was the executive for the pilot study but was

soon replaced by H. H. Stern, who was also on the staff of the Unesco Institute). Robinsohn was cool toward the type of work that the IEA was thinking of doing and was none too happy to have this quite different group of researchers forming an independent unit within his fiefdom. Conditions became so difficult that, at the beginning of the next meeting, in June 1962, Wall had a stormy meeting with Robinsohn about the place of the institute in the new mathematics study, and, as a result, Wall went back to England. Here was a new, large-scale, cooperative venture—suddenly without a chairman. Torsten Husen was appointed chairman, and a Standing Committee, which actually directed the study, was formed and consisted of Husen, Bloom, Mialaret, Robinsohn, and Walker. It was a Pyrrhic victory for Robinsohn, however, because a dinner was held in the Hamburg Ratskeller, to which came the Standing Committee (except Robinsohn) and representatives of each participating country, but no one from the institute. The group prepared an ultimatum that the project could be "under the auspices of" or "in cooperation with" Unesco but not "a project of" Unesco, and funds would not necessarily be administered by Unesco. Husen was commissioned to present the ultimatum.

That very night Foshay received a telephone call from New York telling him that the U.S. Office of Education had promised the money required for the mathematics project. Husen, therefore, had additional ammunition, and Robinsohn accepted the ultimatum. He sent out a memorandum on October 1 to all members of the Unesco Council that formally announced that he had just received word of the grant to meet international costs and that nations would have to meet their own costs. By his intransigence, Robinsohn in fact helped the group of researchers coalesce into a stronger collaborative group than they might have become.

As the mathematics study got under way, a new position was created, that of international coordinator, and a young researcher from the National Foundation for Educational Research, T. Neville Postlethwaite, was appointed to the post. At the same time, Husen

was appointed project director and awarded an honorarium so that he could devote a reasonable amount of time to the mathematics study. With Husen and Postlethwaite as its leaders, the IEA assumed the character it was to have for the next decade. Torsten Husen was the diplomat, and his ability to encourage people to cooperate, to smooth ruffled feelings, and to deal with various funding agencies complemented Postlethwaite's ability to manage the details of the project, to press everyone to a high standard of care with an optimum of speed, and to oversee the technicalities of a complex project so that all went smoothly. These two, perhaps better than any other combination, were able to affect the cooperative nature of the IEA, to make each worker in each national center feel a part of the project, and to expand the IEA from a small group of researchers to a network of over 30 countries during the Six-Subject Survey.

It was in 1963 that another new person was added to the IEA, Gilbert Peaker, one of the most remarkable men I have ever met. A first in mathematics from Cambridge, a marathon runner, and a leading alpinist, he had held various positions in British ministries, from the Cartographic Section of the Colonial Service to the Treasury, where he helped J. M. Keynes, particularly with the rationing schemes of World War II. His main work, however, was in Her Majesty's Inspectorate, where he was concerned with education and with the analyses of the data from educational surveys conducted by the government. Peaker was named sampling expert and later statistical consultant to the IEA. Peaker's first contributions were theoretical and practical: he adumbrated the theory and practical application of drawing probability samples, using a two-stage procedure and a means of calculating design effects; he also made the suggestion that the data from each country be sent to the processing center in four separate packages, each representative of the total so that if one was lost in transit, there could remain some data for each country. He also suggested that the complicated process of calculating sampling errors could be simplified if the four subsamples had been appropriately chosen, as the ranges of

their means could be used to estimate these errors. He also did the first multiple-regression analyses on these data.

With its three coordinators – Cobb (1960-61), H. H. Stern (1961-62), and Postlethwaite (1962-72) – the mathematics study proceeded but not without problems. As it moved along, it began to attract notice in other countries. In New Zealand, the then director of the New Zealand Council for Educational Research, G. Parkyn, wrote to the director of the Unesco Institute and suggested that New Zealand might take part. Stern, the coordinator at the time, replied that there was the possibility of joining in the project on the basis of "associate membership," and this suggestion was confirmed by Husen in 1963, but the notice proved too short, for the data were to be collected that year and there was no provision in the timetable for the Southern Hemisphere school year. At that point, one might say, it was extremely difficult for a system of education to participate if it had not been involved at the first stages of planning. Concern over the seeming exclusivity of the process later led to a more flexible timetable, but that, in turn, was to present other problems. For some of those countries that had joined the mathematics study, the advantages were clear. In Finland, a newly fledged research center, the Institute of Educational Research in Jyväskylä, found that joining the IEA project helped to establish its position with the government and with the research community in that country. As Kimmo Leimu wrote for the East-West Center, "the simultaneously emerging and consolidating IEA-collaboration was an appropriate and important counterpart, being newly conceived and driven by a comparable enthusiasm and devotion to a worth-while cause."

In Belgium, a similar situation occurred; Hotyat had worked alone and without any financial support on the pilot study and the later mathematics study, coding all the data in his home. In the first years of the mathematics study, he was replaced by Gilbert De Landsheere, who was building a department of education research at the University of Liege. Again, the IEA helped his department and its students learn firsthand about survey research and data analysis, and their work

with the IEA helped forge one of the strongest research departments in the French-speaking world. As De Landsheere wrote for the East-West Center, "What did the Belgian national center learn from IEA? In a narrow sense, definite techniques such as multi-stage sampling, multiple regression analysis, score weighing, etc. But it seems to me that even more important is that personal and work relations are created with leading scholars such as B. S. Bloom, R. Thorndike, J. Carroll, T. Husen, G. Peaker, and so many others, with directors of National Foundations for Educational Research, with national IEA project directors and technical officers... The amount of information, help, and advice thus made possible is invaluable."

As the mathematics study moved through the testing and data recording phases in 1964, some consideration arose for the future of the IEA and, in particular, its relationship with the Unesco Institute of Education. Through the mathematics study, the IEA could be considered a group of 12 researchers who agreed to collaborate on a project; some came from universities, some from research institutes partially or totally supported by the government. But they acted as individuals sharing responsibility for test construction, research design, data processing, and data analysis. They met as a committee of the whole at least annually and appointed a trusted subgroup to act as their surrogates in between times. This collegial relationship was effective, but if the IEA were to expand, and many thought it should, it needed somehow to modify that collegial relationship and create an organizational structure. Such was to become particularly the case when the areas that were researched expanded. At the October 1964 meeting of the Standing Committee, it was agreed that the subjects to be explored were science, literature, civics, mother tongue, and modern languages. At the same time, a long list of countries had written about joining, and the committee had to come up with a set of criteria for membership. Other problems concerned finances; the coordinators were never aware of how much money had been given the mathematics study or how it was allocated.

The report of the committee was brought forward at the next full meeting, held in Chicago in 1965, but the attention of the group was riveted on the data from the mathematics study. There were plenary sessions interrupting group meetings on the various hypotheses and meetings of the Standing Committee and the editorial committee responsible for shaping the two-volume report. The meeting was hampered by the University of Chicago computer, which could not meet the too-high expectations of the researchers and which presented the group with many problems, the most famous of which was its finding that there were nine sexes in Sweden, a result reported by a straight-faced Robert Thorndike at the subsequent meeting of the American Educational Research Association (AERA). Many of the problems could be corrected, but, in some cases, the data were so erroneous that certain countries had to be dropped from some of the analyses.

It was at this meeting that Gilbert Peaker reached the conclusion that regression analyses might be the most useful way of analyzing the data, although he was well aware, as David Walker said, of "the snags that can arise when the independent variables are in fact highly correlated" (i.e., the problem of multicollinearity). The Peaker regression analysis used only 25 variables. On this as on some other issues, there were sharp disagreements, but these were always voiced within a friendly atmosphere, and the group's spirit of cooperation was high. In fact, it might be argued that it was made higher by adversity, by a keen sense that something exciting was going to emerge, and by the fact that these were like-minded people working on a common problem that engaged their intellectual interest.

### **Second Phase: The Stockholm Empire**

The second stage of the IEA dynasty began in 1965 and moved certain people into center stage. Some of them had been highly active during the mathematics study, but they came into their own with the Six-Subject Survey. Torsten Husen, Neville Postlethwaite, Benjamin



Bloom, and Gilbert Peaker remained, but others, including Gilbert De Landsheere and Douglas Pidgeon, became central to the survey. To them, of course, must be added the subject committee chairs and many others.

Beginning as early as 1964, the IEA faced the problem of its legal status. It had been fairly easy to exist legally under the wing of the Unesco Institute, although living there had its difficult moments. It was also fairly convenient for the U.S. Office of Education to provide funds to institutions such as the University of Chicago to carry out the work, but that seemed to signify American domination, and other agencies might want to invest funds. In his paper for the East-West Center, De Landsheere chronicles the sequence of events leading up to incorporation in 1966:

It was the Ford Foundation who raised the problem of the *personnalite juridique* when the Foundation made it a condition for further grants:

- a. that IEA should possess the *personnalite juridique*
- b. that IEA should produce a document certifying that no taxes would have to be paid on grants.

All developed countries have made provisions for incorporation of national associations of all kinds, but not for international associations. That is why IEA had to ask for the help of international law specialists (from the UNESCO Paris legal section) to try to discover where in the world appropriate and advantageous status could be found. At the end of long and difficult inquiries, it appeared that Belgium was probably the most hospitable country in this regard. After World War I, King Albert indeed dedicated much energy for the development of scientific research in Belgium, for he felt that research was a key not only to the reconstruction of the devastated country but also for its future development: to that effect,

and among other things, he created the National Foundation for Scientific Research, and issued a so called frame law for incorporation of international scientific associations wanting to have their seats in Belgium. In fact, this law combines the usual statutes of national non-profit organizations with the optimum condition for the development of international scientific associations.

The association was officially incorporated (in 1966) by a decree, signed by King Baudouin.

Under the statutes, the membership of the IEA was to comprise mainly institutions, although there could be a small number of individual members. The institutions were to be restricted to one per country, but such a restriction brought a number of problems. The first was a definition of a country. From the very beginning, both England and Scotland were members of the IEA; they were politically one, but educationally two. Another example was Belgium, which in 1966 had established a French-speaking system and a Flemish-speaking system. In 1967, it was resolved that, instead of countries, the IEA should consider systems of education. Another problem came with the United States, which had two universities represented – one receiving a grant for the mathematics study, another for the Six-Subject Survey. Again, this was resolved by allowing for two institutions in one country. The problem, by the way, has not disappeared; currently, Canada, with distinct educational entities in each province, is one example, as is the case with the Federal Republic of Germany. Despite the 1967 resolution, the definitions of a country, a political entity, and an educational system remain difficulties.

Another aspect of the organizational problem came from the relationship with the Unesco Institute. Although Robinsohn had been replaced in 1964 by Gustaf Ogren, who was comparatively friendly toward the IEA, the building was simply unable to house two organizations. When the constitution was written, Hamburg was mentioned

as the headquarters, but the space pressures mounted and there were no computing facilities. In 1968, Husen initiated discussions with Stockholm University and received an offer of 120 hours of free computer time for the IEA and good office facilities as a part of his personal chair. T. Kobayashi, the new director of the Unesco Institute, indicated that the accommodations simply would not permit the expansion of either the IEA or the institute, not to mention the difficulties in housing two organizations with different objectives and different working methods. On July 11, 1969, IEA headquarters in Hamburg closed and on July 14 reopened in Wenner-Gren Center in Stockholm as a fully operational unit. The person responsible for so smooth a transition was Ann Irwin.

The financial problems of the IEA continued during those years, but there were successful applications to the U.S. Office of Education; the Ford, Volkswagen, and Leverhulme Foundations; and the Bank of Sweden Tercentenary Fund. Near the end of the Six-Subject Survey, there was also a successful application to the Spencer Foundation to establish a series of fellowships for various researchers to work with the IEA data, which Thomas James, then president of the foundation, told me he considers one of its best investments, in that the fellows have gone on to positions of leadership around the world.

All of these issues were, perhaps, less central to this phase of the IEA's life than were the conceptual and managerial aspects of conducting a massive survey of achievement in six subjects: science, reading comprehension, literature, civic education, French as a foreign language, and English as a foreign language. In addition, there came the addition of new participating countries: Chile, Iran, Poland, and Thailand were admitted in 1967; New Zealand, India, Hungary, and Romania in 1968. The motives for joining were many. New Zealand wanted to test its new science curricula by establishing some baseline data on the current curriculum and briefly considered entering as a "seventh state" of Australia, but when the Department of Education found out that there were other subjects being tested, it decided to join on its own. Chile

was moving to a new system of university entrance examinations and also needed reliable information on student achievement. Thailand wanted to evaluate its programs, but, more important, it wanted to have contact with other researchers. These countries and others frankly wanted to learn about large-scale evaluation of educational achievement, and they wanted to go to school with the best faculty. The IEA at that time clearly had the "best faculty," particularly in curriculum analysis and test construction, with Pidgeon, Thorndike, Carroll, and especially Bloom, whose taxonomy was coming into its own around the world. In addition, there was Peaker for sampling and he, Thorndike, Richard Wolf, and Walker for data analysis and interpretation, not to mention others, such as Lee Cronbach and James Coleman, who occasionally gave advice. To a great extent, however, it was Benjamin Bloom who was the intellectual leader of the mathematics study and the first phase of the Six-Subject Survey and who saw in the IEA its greatest potential, the training of a cadre of educational planners and evaluators around the world. Both he and Peaker dominated the studies.

The reason for arguing Bloom's centrality comes from the fact that he laid the plans for the mathematics volume, dividing it by hypotheses. More important, in each of the six-subject studies, the approach of adapting the cognitive taxonomy (which is informally known as Bloom's) and the affective taxonomy as well to the subject matter in order to define national and international objectives enhanced and strengthened the earlier concept of opportunity to learn. In most of the subjects, one could obtain an estimate of the overall national opportunity to learn or the intended curriculum, then the teachers' estimates of opportunity based on specific items or the implemented curriculum, and finally the actual student achievement scores. This approach was clear in the science, civics, and second-language studies and to a lesser extent in the literature study, and it has come to be the operative approach in the second mathematics, science, and written composition studies. Beyond this influence was a clear sense, deriving from the tax-

onomy in the affective domain, that an important aspect of the IEA studies should be the inclusion of attitudinal items and the conception of these as educational outcomes.

At the same time, the inclusion of attitudinal items led to many lively debates. In the Literature Study, there was proposed a set of items dealing with censorship. These items were objected to by the Iranian center and one or two others, and so they were dropped. In the Civic Education Study, however, attitudinal questions became a central issue. Even the cognitive items appeared to be value laden, and many were accused of being biased in favor of the U.S. system. It was clear, for example, that such terms as "democracy" and "social welfare" had different meanings in different countries. There were lengthy discussions of this issue, and at one point during the debate the Civic Education Study was on the brink of collapse, but a sufficient number of countries wanted to study the field. Douglas Pidgeon summarized the outcome of the debate; that the study would continue internationally, that the instruments would be revised to remove bias, and that many measures would be treated as attitude questionnaires rather than as cognitive items. Nonetheless, some countries, such as Chile, dropped out of the Civic Education Study because of its sense of a remaining bias.

Other studies had equal problems. The foreign language committees wanted measures of oral performance, but some countries simply did not have the equipment needed to record oral performance. In addition there were grave issues about the reliability of scoring. In literature, there was a strong bias against multiple-choice items, and a feasibility study comparing them with open-ended responses had to be undertaken to prove their value. In reading, a word-knowledge test was developed to give an estimate of basic verbal proficiency, but it became drastically more difficult when translated into certain languages. In the science study, a general measure of the understanding of science was developed, which also proved problematic in interpretation, for it too was value laden.

Despite all these difficulties, the studies were carried out with the testing spread over 2 years so as to ease the burden on the schools and to allow those more complex subjects more time for tryout.

At the same time, two other activities began to emerge: model building and training. Following the mathematics study, it appeared that an "input-output" model was appropriate. So, as a part of the IEA's activities, but funded separately, two conferences were held, one in Hamburg and one outside New York City, to develop a strong input-process-output model, focusing particularly on the independent variables—personal, environmental, and national (Super, 1969). The effect of these conferences on other comparative educational studies is probably mixed—if, indeed, the reports are known. This attempt at model building was partially followed through in the development of the regression models used in the Six-Subject Survey, which sought to analyze both individual student performance and school performance. It was here that Peaker's intellectual leadership was evident, for his model with the famous "yacht handicap," which linked students to the ecological niches of schools, provided the conceptual framework for each of the volumes.

In a yacht race, the performance of a captain, crew, and boat is judged not by who crosses the finish line first but by the amount of time taken to sail the distance after the size of the yacht and its sails are taken into account. Each yacht, therefore, is awarded a handicap depending on its waterline length, its sail area, and several other factors, and the actual time taken is adjusted for this handicap before the performance is assessed. A similar matter occurs in assessing methods. One can consider the performance of individual schools, but such consideration is not as meaningful as a consideration of what the schools do with the students they receive and therefore with the degree of change they render with the people they have. In part, the school receives students already determined by the community—its socioeconomic level and its interrelated level of "culture" are two of the major forces in defining that community. These measures would define for the school its con-

textual setting, and the effectiveness of the school can then be seen in terms of what has been achieved after allowances for the context have been made.

In his paper for the East-West Center, Keeves suggests that the model was imperfectly followed in the analyses, and possibly for the best. From the attempt at model building, there also came the realization that survey questionnaire data did not adequately capture what went on in the classroom. A review of research by Barak Rosenshine was commissioned, and plans for some sort of classroom process or classroom environment study to follow the Six-Subject Survey were discussed in June 1972.

The training function was, perhaps, more important than the model building. To be sure, there was on-the-job training for the national technical officers, as they were then called. Usually either Postlethwaite, Pidgeon, or both visited each of the centers for several days. At times their visits were more than training. One center suffered a change in personnel, and the outgoing director destroyed all the IEA documents (or took them with him), and Postlethwaite had to reconstruct them all and explain the whole project from beginning to end for the unfortunate replacement.

But there was more formal training as well. In 1968, John Carroll gave a seminar on research and educational process. This was repeated in France in French (1970) and in Germany in English (1971). In 1971, at the time that the first results for the Six-Subject Survey were coming off the computer, a seminar, which included 123 participants from 23 countries, was held in Granna, Sweden. The seminar was directed by Bloom, managed by Postlethwaite, and included several of the Subject Committee members on its faculty. The participants were teams from each country that were expected to work together on curriculum development and evaluation. Most of the teams came from Latin America, Asia, and Africa, and they were treated to a variety of instructors, most notably Ralph Tyler. From Granna there emerged

two results: a group of informed researchers who were to form the nucleus for the next round of IEA studies and the idea for an organization parallel to the IEA called the International Curriculum Organization. Bloom was the progenitor of this idea, which never took hold, in part because, although Husen tried to get the International Institute for Educational Planning (HEP) to back the idea, it ran afoul of Unesco, some of whose staff saw it as an intrusion on their domain.

In 1971, as the data from the first three of the Six-Subject Studies began to come to Stockholm, several authors spent the summer there working on their volumes when they were not lecturing at Granna. Unfortunately, however, most of the analyses were not performed until later in the year, and the actual writing had to wait until 1972. In August 1972, Postlethwaite left for Paris, and the HEP and Husen appointed Roy Phillips, the national research coordinator for New Zealand, to take over the editing of the volumes and the final coordination of the project. This he did capably for 2 years, aided by various Spencer fellows, particularly Leigh Burnstein and Jack Schwille, who with Richard Noonan performed many of the analyses for the authors of the second-stage volumes.

The results of the science, reading, and literature studies were published in 1973, and, in the fall of that year, there was held what might be seen as the culminating event of this stage of the life of the IEA, the Harvard-IEA Seminar, the results of which were published in 1974. Paul Ylvisaker, incoming dean of the Harvard Graduate School of Education, wanted to do something that would help reestablish Harvard internationally in the field of education and offered to host a 1-week seminar for IEA members and others to discuss the results and implications of the study. Although the foreign language, civics, and national case-study volumes were not out, there were reports on the preliminary results. In many ways, the Harvard-IEA seminar was a success, not least in that it encouraged many people that the IEA should continue.



### **Phase Three: The Satrapies**

By 1973, the first of the oil crises had hit. One respondent at the Harvard-IEA Seminar, the Boston school superintendent, said that it was difficult to evaluate the results of the studies and their implications because his concern was whether he would have enough oil to heat the schools that year. Undoubtedly this concern was shared by his counterparts in other countries. Nonetheless, there was a desire to keep the IEA fires burning. To be sure, there was another year's work on finishing up the civics and language studies and to write the technical and summary volumes, and there were 2 years to go on the Spencer fellowships. But many of those who had been working with the IEA for a number of years were retiring or simply involved with other activities, and all were tired and weary after 7-10 years of intensive work.

At a meeting in Stockholm in 1972, there was some discussion of various new studies: the four for which proposals were drafted were a repeat of the mathematics study, a study of early reading, a study of preprimary education, and a study of classroom processes or class environments, with a particular emphasis on mastery learning to test whether it could work effectively in developing nations. From among these various proposals, the one that appeared most feasible was a repetition of the mathematics study. Roy Phillips, who was finishing his tour of duty in Stockholm and about to return to Wellington, was the strongest proponent of such a study, and an exploratory group was formed. This group wanted to include a longitudinal component to the study in order to look at both growth and teaching practices. Such a proposal was advanced at the 1974 meeting in St. Andrews, Scotland, but the conceptualization did not appeal to the council. Matters were held in abeyance until a meeting at the Max Planck Institute in Berlin in 1975, which was held to discuss the results of the Six-Subject Surveys and in general to discuss the effect of the IEA on both industrialized and developing nations. At that meeting, which to many

appeared to be the swan song of the IEA, I was approached by Husen and Phillips to see if I knew of anyone who might take charge of a mathematics study. When I returned to the University of Illinois, I spoke to Kenneth Travers, who, although not involved with the first IEA mathematics survey, had some experience in survey research. He agreed to investigate the matter and together with Phillips secured the funds to have a meeting in May 1976 to bring together representatives of other countries who had international experience in mathematics and survey research, particularly in the Six-Subject Survey. At the May meeting, the group discussed a number of issues related to a second mathematics survey and encouraged Travers to work on a proposal, which he did with a smaller group that met in St. Andrews in June. The proposal was supported by the Ford Foundation, the Spencer Foundation, and the National Institute of Education (U.S.), and the study was launched in 1977.

Meanwhile, various people were still pushing for a classroom environment study. The General Assembly met in Tokyo in early 1978, where a group that included Arieh Lewy of Israel and Nathaniel Gage of the United States brought forward a proposal for the classroom environment study. By the end of the meeting, this proposal was authorized to begin, and funds were sought for the initial coordinating center to work in Tel Aviv. The proposal from Travers and his group for the second mathematics study was also presented (in Tokyo), and, although some eyed it with skepticism, there was much encouragement to proceed. The international coordinating center was established in Wellington and the technical offices in Urbana, Illinois.

Also at the Tokyo meeting, Postlethwaite was asked to be chairman of the IEA with Torsten Husen taking the role of president. Postlethwaite, who had left the HEP to accept a chair in comparative education at the University of Hamburg, agreed, and the IEA headquarters was split between Hamburg and Stockholm, which kept a skeleton office for routine inquiries. At the same time, there was discussion about various other aspects of the IEA's work: whether there should be new

studies, the role of training, and the general organization of the association.

By the 1979 meeting in Paris, there began to emerge the current structure of the IEA as the new studies took on a life of their own. A coordinating center for the Classroom Environment Study was established at the Ontario Institute for Studies in Education, with Doris Ryan as international coordinator. Other groups were encouraged to formulate plans for new studies; one in science to be directed by John Keeves and one in written composition to be directed by me. These proposals were ratified in 1980 at the General Assembly meeting in Jyvaskyla, Finland. By the end of 1980, then, headquarters for IEA studies were in Australia, Canada, New Zealand, and the United States. Plans for other studies were at the same time beginning to be formulated: in the language of the home and the language of the school, in preprimary education, and in the transition from school to work. Today, these studies are at various initial stages, and new ones in the uses of computers, in reading and literacy, and in values are being initiated. If funded, these studies will spread the headquarters of the IEA even farther. The reasons for decentralization included the pragmatic ones that each study secure its own international funds and that there be no full-time salaried employees of the IEA and the operational one that it seemed better to let specialized studies have independent lives.

Such decentralization has been accompanied by plans for centralization and greater structure of the IEA as more and more educational systems—especially from so-called developing countries—have come to be involved. During the period 1982-84, representatives of the World Bank wished to create an international fund for educational research in developing countries and for some of the funding to be used for facilitating developing countries' participation in IEA projects. The plan has been frozen within the Bank. The Bank remains deeply interested in the IEA and its work. Following this there has been established a plan for a centralized data-processing center in the Netherlands, to serve all the studies. At the same time, the General

Assembly has worked to modify and develop its bylaws and other operating committees – on publications, on technical support, on policy and procedures, and on dissemination – so as to allow a decentralized cooperative venture to function as smoothly as did the former centralized venture of the Six-Subject Survey. Such an effort has not been easy, to a great extent because the financial resources have not been available to allow for frequent meetings and visits to the various centers. Correspondence is a chancy mechanism, and there is always the mailroom clerk who will take the responsibility for determining that a data tape should be sent by sea mail from Europe to New Zealand, thus setting data processing back 5 months. Although an imperfect solution in many ways, the satrapies of the IEA function about as well as could be expected, although not as well as could be hoped.

At present, each study depends on the collegiality of the individuals working on it, as does the General Assembly. There have been many periods of friction, to be sure. The operating committees mentioned above are advisory; there are few sanctions either on a particular study or on a participating center that fails its obligations in a study. The IEA still depends on a cooperative spirit, but that spirit has its limits, and there remains a tension between the desire for free inquiry and that for rigorous standards of quality control. Although the IEA no longer represents a single research paradigm, it seeks to insist on high-quality work within the paradigm selected. It seeks also consensus among participating systems. As the IEA becomes larger, consensus is harder to attain. The organization now involves a process of socialization rather than a strict hierarchical structure. Whether this approach will be successful for the long run remains to be seen.

### **A Retrospective Conclusion**

A memoirist is, to a certain extent, a moralizer, and many biographies are exempla, a tradition dating to Plutarch. In chronicling the unfin-

ished life of the IEA and its emergence from a confederacy of colleagues to an empire and to its present state of a collection of satrapies, I see certain lessons to be learned. The first is the importance of people, the second the importance of communication, and the third the importance of tempering a model with reality.

During the course of this memoir, I have singled out a few of the leaders of the IEA, but to do so is, perhaps, to do disservice to the enterprise. From the papers of the East-West Center project, a few quotations indicate that the success of the IEA has rested more on the many than on the few.

Postlethwaite: "If I were to pick any one factor which was primarily responsible for the success it was the dedication of a handful of workaholics who were in key positions in the project."

De Landsheere: "I would type most of the correspondence myself... and pay for the post-stamps... These things may seem trivial, but when you teach many hours a week, do your own (heavy) research,... have to work with students on their papers and thesis, take part in several national committees, and have some international activity, weeks and months are really very short."

Renwick/Phillips: "Most of the collation exercises were undertaken within the unit. The sight of several highly paid curriculum officers spending their lunch hours walking around a table of papers was not an unusual sight."

Rodriguez: "This [visits to schools] was a very hard job, some researchers had to reach the schools on horseback since no other means of transportation existed."

It has been the custom of the IEA that, with few exceptions, no one is paid for international work such as serving as a member of the General Assembly or a project committee. Such work is "stolen" from one's workdays or vacations. As these testimonies suggest, that attitude permeates the studies down to workers at every level. In addition, most of the people find it intellectually stimulating to work on a

project, and from participation in the IEA have come strong friendships. As the IEA has become a set of parallel projects, it has been the project more than the IEA that has the *esprit de corps*; but the General Assembly meetings are times of both hard work and good fellowship. The second lesson, communication, is perhaps less positive than the first. As the memoir indicates, the meetings have been held primarily in English, with some French and German being used. This fact has itself created problems, for many first-rate researchers have a reading knowledge of English, but their speaking knowledge is not as good as it might be. As a result, many of the meetings are hampered by poor communication, and many of the documents are produced slowly. Each study has sought to prevent English as the medium from turning into Anglo-Saxon dominance of the studies, although it is clear that most of the impetus has come from those systems.

Communication is not simply a matter of language, however, but of culture in its broadest sense. In the various centers, there were differing styles of work, differing senses of what might constitute good execution of a sample or good methods of analyzing data, and certainly differing bureaucratic structures. In some centers, the director was the main worker; in others, simply a paper shuffler. At the international center, one never quite knew which one was dealing with. One never quite knew, when a national center did not respond, whether the original communication was lost, the institute was on strike, or there was simply no money for international postage (all three reasons have come up from time to time).

As the memoir indicates, the early projects were marked by frequent meetings. As the IEA expanded and as funds, particularly international funds, have become less easy to come by, meetings have become less frequent for some studies. They also tend to become less frequent at the analysis phase of a project, which is as crucial to the life of the project as is the conceptual phase. As a result, more communication has to be done by mail, and often there can be misunderstandings. A famous

example occurred during the Six-Subject Survey when the data processing was done in New York and the analyses in Stockholm. The first correlations arrived in Stockholm, and all were surprised to find that interest in science and achievement in science were negatively correlated in every country. The same was true for interest in literature and achievement in literature. Four of us spent a day and a half trying to find an adequate interpretation for the result. A telephone call to New York finally produced the reason: the scores on the interest measures were reversed so that a low score meant high interest. Other occasions have not been so humorous or so easily resolvable.

The IEA has generally striven to lessen problems in communication, and its members all try to be forbearing and diplomatic, but occasionally they fail, particularly when information is relayed through third parties, as often happens. The organization has also striven to increase communication between psychometricians and subject-matter specialists, between researchers and policymakers, and between researchers and teachers—all of this across national and cultural boundaries. It is indeed a difficult task, but one well worth the struggle.

The third lesson arises from my sense of how the IEA has operated at its best intellectually. When the Foshay proposal was inaugurated, the hope was expressed that there some cognitive universals might emerge. What did emerge was the importance of the curriculum and opportunity to learn. In the first mathematics survey, an input-output model determined the structure of the study, but the most important finding in many eyes dealt with the notion of the "yield" of an educational system. In the Six-Subject Survey, a comparison of home effects and school effects dominated the study, but perhaps the most important findings related to the "unintended" attitudinal outcomes of schooling in science, civic education, and literature. In all these studies, the model that one intended to validate empirically proved, perhaps, to be less important than certain other findings. In part, this might be the result of the length of time of an IEA study.

To a greater extent, however, I believe this results from what has emerged as the prevailing metaphor of the IEA, that it sees the world as an educational laboratory. There has emerged a natural variation of school and social systems, each one contextually embedded. To look at the effectiveness of any one system, one needs to get beyond it. That is why there are comparative surveys. One can compare such phenomena as size of class, age of beginning school, or pattern of teacher training without disrupting the entrenched system of any one country. An IEA study at its best includes what Gilbert Peaker called "hunting and fishing." To hunt in research is to build a tight model and test it. To fish is to build a loose model and see what turns up.

In as complex a laboratory as the IEA's, where context is important and one is always a little unsure about the precise nature of anyone else's educational system, some models based on research in one country maybe inapplicable in certain contexts. Being model-less is also dangerous because one may be simply surrounded by data and not able to maneuver through it. The IEA has consistently included independent variables gathered according to well-formulated models arrived at consensually and tested internationally. It has also included what Peaker called "favorite sons," those variables that look like a good bet because they make intuitive sense in one country, yet the question remains whether the intuition is culture specific or universal. Increasingly, we have come to see that culture modifies the universal and that the examination of cultures has become increasingly important.

Model building in the IEA has led to increasingly complex models compared with those proposed in the early studies. The IEA researchers have come to see that schools and educational systems are deeply embedded in a matrix of economic, political, and cultural traditions; that times change but at different rates for different parts of the world; and that schools, subject matters, and people have both universal and idiosyncratic characters. Survey research can tend to flatten



out the idiosyncrasies, but the IEA has tried to make them remain rounded.

Through the course of its history, the IEA has changed the consciousness of nearly everyone who has participated in its various studies. It has, above all, sobered their heady belief that there is one best way to educate the young. Although the IEA has the aspects of an empire, it does not have an imperialistic mentality.

### References

- Foshay, A. W. (Ed.). (1962). *Educational Achievements of 13 Year Olds in Twelve Countries*. Hamburg; Unesco Institute of Education.
- Super, D. E. (Ed.). (1969). *Toward a Cross-national Model of Educational Achievement in a National Economy*. New York: Teachers College Press.

## CHAPTER 24

# Experiencing the World as an Educational Laboratory

*Alejandro Tiana Ferrer*

When I became acquainted with the IEA in 1990, it was already a mature organisation. Yet in the last twenty years, I have seen it grow further and have accompanied it in that process. There are perceptible differences between that earlier association I first knew and what it has become today. In response to the questions posed by the editors, I will use these pages to take stock of my personal and professional experience over two decades in international research on the evaluation of education systems.

### **My Encounter with the IEA and its People**

In 1989, I joined the Spanish Ministry for Education and Science as director of the Educational Research and Documentation Centre (Centro de Investigación y Documentación Educativa - CIDE). I had come from the National Distance Teaching University (Universidad Nacional de Educación a Distancia - UNED), where I taught at the Department of History of Education and Comparative Education and had run the Educational Sciences Institute. My academic background had little to do with evaluation, a subject on which I had barely touched previously.

My mandate as director of the CIDE, however, included reinforcing evaluation activities in the Spanish education system, then in their infancy. While in this post, I had the opportunity to supervise Spain's participation in International Assessment of Educational Progress (IAEP) studies and in the IEA's Reading Literacy Study, as well as to publish and disseminate evaluation reports on the experimental

reform of secondary education which took place in the Eighties.

Though I did not consider myself an evaluation expert, my knowledge of statistics and my dedication to education policy combined to arouse my interest in what I considered (correctly, I now see) would be a growing area in the years to come. My first foray into this field revealed the potential that evaluation offered for the effective monitoring of education processes and results, as well as the risks that its misuse could entail. I discovered at that time that evaluation was becoming a way to support both the making and the explaining of policy, in education and in other areas.

Thus it was that I first came into contact with the IEA in 1990. Spain had formally joined the association a short while earlier, and the CIDE was the body that represented it. As the Centre's director, I had to attend general assemblies and supervise the work of national coordinators in diverse studies. We opted to take part in TIMSS 1993 and to support Catalonia's involvement in the Early Childhood Education Study. Spain became an active member of the IEA, explicitly supported by the Ministry for Education and Science.

Spain was not an exceptional case at that time. It is true that, up to that point, not many countries were engaged in regular and systematic education system evaluation. However, there was a growing and tangible interest in these activities, which had led to an increasing number of international studies, and like us, many countries were already organising their own national evaluation systems. In Spain, that mission was given to the National Quality and Evaluation Institute (Instituto Nacional de Calidad y Evaluación - INCE) by the Education Act of 1990. The new institute began to operate in 1993 and I was its director until 1996.

Over those six years, I was actively involved in the IEA. I attended all its general assemblies, was on the Standing Committee between 1992 and 1994 and, together with my CIDE colleagues, organised the 1993 general assembly in El Escorial, near Madrid. Spain became an active and engaged IEA member and was recognised as such.

For me personally, it was a period rich in experiences, as I was able to become acquainted first hand with the new developments taking place. I had the chance to observe international trends from a privileged viewpoint, became acquainted with prestigious researchers whom I had previously seen only in print and discovered that many of them, as well as very valuable professionals, were also excellent people. I had interesting conversations around the world with the likes of Gilbert de Landsheere, Tjeerd Plomp, Seamus Hegarty, Richard Wolff, David Robitaille, Al Beaton and many others from whom I learned so much.

Those researchers came from different places and traditions other than mine. My own academic work had above all covered the history of contemporary education systems and comparative education in European countries, and here I was faced with teachers and researchers from all over the world who had conducted rigorously quantitative empirical research, using advanced and sophisticated instruments with which I was barely familiar. It was a true challenge and one that required a great deal of study on my part, not to mention a rapid immersion in specialised English.

In the 1990s, when I first became involved in the IEA, the organisation had taken on many people since its creation thirty years earlier. People like Torsten Husén rarely attended its general assemblies by that stage, while Neville Postlethwaite still took part, though less often than before. Yet I never had the chance to meet some of the founding fathers about whom I had heard so much. Apart from them, however, there were also many new researchers, some of whom were to take on major responsibilities in the years to come. The IEA was an organisation in the process of renewal, where researchers and political leaders still in their youth could be found working alongside established academics on the point of retirement.

Of all these individuals, I have particularly fond memories of Tjeerd Plomp, with whom I had a magnificent personal relationship. Shortly

before my arrival, he had been elected IEA chair and was enjoying his first term at the head of the organisation. I witnessed the hard work he put into giving the organisation new impetus, the difficulties he had to confront and his constant determination to put his weight behind the common cause. He placed his trust in me by proposing me for the Standing Committee in 1992, just two years after my arrival, and I was finally to take over from him as chair in 1999.

To sum up, I will say that my arrival at the IEA broadened my professional horizons, put me in contact with the international research community and brought me some very good friends. It demanded a learning process that was as fast as it was intense – based on the experience and implementation of international projects – which would otherwise have taken me a great deal longer.

### **The General Philosophy of IEA: the World as an Educational Laboratory**

One of the aspects of IEA that most attracted me was its idea of the world as an educational laboratory, a metaphor first introduced by Torsten Husen. This statement was more an aspiration or a pledge than a reality, but that probably was one of the main reasons for its attractiveness.

The conception of the world as an educational laboratory was based in what constituted, in my opinion, the main novelty introduced by the IEA in the sphere of educational evaluation: to draw attention to the need to assess students' achievement. Although this has since become common practice, at the time it was a new idea. At a time when a large part of research was geared towards studying and evaluating the resources available, or the conditions in which teaching and learning processes were taking place, the IEA held that the evaluation of an education system ought to be based on a rigorous and objective assessment of the results students achieved. It was with this aim in mind that it designed and set up its earliest projects.

However, the IEA researchers did not merely restrict themselves to the measurement of student achievement in different areas of the curriculum in the countries taking part in the surveys. They also sought to explain the results obtained, analysing a series of curriculum and background variables which they considered could affect performance. The identification of some of the main factors behind the achievement of better results would enable decisions to be taken that could improve education, which was, after all, their chief concern.

It has to be admitted that these ideas were not exclusive to that particular group of researchers. In the Sixties, when the IEA was embarking on its early work, such an approach had already become fairly widespread. For me, the main novelty which the association injected into the debate, and which was undoubtedly one of its major achievements, was its conception of the world as an educational laboratory. This catchphrase, which gained currency in the Sixties, was evidence of the association's desire to adopt a truly international and comparative perspective with which to conduct its evaluation of academic performance.

The empirical work the researchers had undertaken in their respective institutions made them aware of the difficulties and limitations, both technical and ethical, of attempting to apply strictly experimental methods to the world of education. Although small-scale experiments can usually be conducted in education, studying change in a broader context is a different matter, particularly where it affects education systems as a whole. It is neither possible, nor indeed advisable to introduce major changes into an education system for the simple purpose of evaluating the consequences. A different approach is needed.

The methodology developed by the IEA was based on comparing the outcomes obtained by students in different education systems, monitoring as far as possible the major variables influencing results. This comparative view would enable the association to analyse and evaluate the influence of the factors identified, irrespective of the specific

circumstances of individual countries. The *natural diversity* that existed between countries in aspects such as the organisation of the education system, teacher training or the curriculum taught – to name but a few of the potential differences – would mean their effects on results could be analysed and evaluated without needing to introduce complicated and risky change to gauge their influence. Thus, international evaluative surveys would become a powerful empirical research tool in assessing performance and the factors associated with it, from a comparative standpoint and for the benefit of educational improvement.

In order to compile the data required on the main performance-related variables, the IEA began to use different types of questionnaires aimed at those who played a key role in education, in combination with the tests that constituted the main focus of its work. The students examined, their respective teachers and school principals were all provided with specific questionnaires, which were sometimes accompanied by questionnaires for the pupils' families. The aim of this set of instruments was to collect data on how schools operated and how teaching and learning processes took place within them.

It was on the basis of this thinking that the IEA began to develop its first international surveys in the Sixties. All those projects shared three distinctive features: 1) they focused on the evaluation of student achievements in different areas of the curriculum, and also explored the problems posed by measurement of these; 2) they adopted an international perspective with the aim of conducting comparative analysis based on the differences found; 3) they used quantitative techniques taken from the most advanced psychometric and educational research, as the best means of obtaining objective and reliable results. The idea of considering the world as an educational laboratory thus took on the shape of a research design that was totally new in the mid-20th century and provided the inspiration for many subsequent international studies. Even though not all their promises have been kept, the idea proved attractive and it can be considered promising still today.

## **IEA's Contribution to International Performance Assessment**

The implementation of the first international studies conducted by the IEA in the Sixties and Seventies required the development of a methodology that has been the subject of continual fine-tuning ever since. Carrying out large-scale comparative studies involving a great many countries poses a series of methodological, technical, linguistic and cultural problems that can neither be ignored nor avoided. The ability to respond to these questions and problems has been one of the major successes of IEA throughout its history and in my opinion has been its most outstanding contribution to the work of assessing educational achievement.

### ***a) Linking Performance Tests with the Curriculum***

The first problem posed by this type of study involved ensuring sufficient coherence between the tests implemented and the aims set for the different educational stages. In other words, if these studies are to help improve education, they must refer to the real circumstances being analysed. Specifically, these circumstances revolve around the school curriculum at a given educational level, which may take on very different forms and characteristics from one country to another. Consequently, the tests used must be linked to the curriculum, in a context of international diversity.

Not all international studies are equally related to the curriculum for the stage studied. For instance, the OECD's PISA project places more emphasis on the acquisition of the skills or competencies young people will need to cope with adult life rather than on the curriculum taught in schools. But the IEA has always aimed to provide measures to improve education systems and, consequently, has focused on the teaching and learning processes; thus most important to IEA is the manner in which the curriculum is designed and put into practice in the classroom.



To assert that there ought to be a link between the tests conducted and the curriculum is a simple matter, although actually achieving this effectively is more complicated. To achieve coherence between the two elements, the IEA has developed its well-known distinction of three levels of the curriculum: the prescribed or intended, the implemented and the attained curriculum, which has proved to be of great benefit to international evaluation studies and has become a common referent in the field. Analysis and comparison of the implementation of the curriculum at these three levels allows interesting conclusions to be drawn about how education systems really work and enables the results actually achieved to be considered in the context of broader objectives.

This concern with the implementation of the curriculum at school has also raised questions about the students' real potential to attain the learning aims set out for them. This consideration has in turn led to the suggestive concept of *opportunity to learn*, based on an analysis of the opportunity students have had to acquire certain knowledge, as a result of the education they have received from teachers and through didactic interaction. Comparison of this factor with the results permits a deeper connection to be established between the curriculum and learning.

### ***b) Designing Unbiased Evaluation Instruments***

One of the main problems that arises in the attempt to assess educational achievement in an international context is to ensure that the study is equally fair for all participating countries. Effectively, ensuring that the tests are equal for all concerned implies using instruments that do not favour countries with particular cultural specificities or languages over others. IEA has always made a sincere effort to ensure fairness in conducting evaluation studies, and has developed very interesting instruments for doing so. In my opinion, this has been an outstanding contribution and one that should be emphasized.

Contrary to popular belief, the above notion does not mean that international studies must be restricted to the evaluation of elements that are common to the curricula of all participating countries. To proceed in this manner would significantly diminish the purpose of the evaluation, given the diversity of content and approaches that must occur in this context. In fact, the aim is to ensure that the tests do not only cover the curricula of individual countries, thereby giving them an advantage over others, but that they also include content that is present or absent to the same extent in the different curricula, thereby ensuring equal terms for all.

But whereas analysis of the underlying cultural elements can be achieved relatively simply through the use of cooperative procedures to construct the tests, dealing with the linguistic obstacles to these issues is more complex and poses greater technical difficulties. Over time, however, IEA has developed a set of procedures to carry out this type of study, and these yield satisfactory results where they are properly applied:

- In the first place, international studies require careful comparative curricular analysis. This is a necessary preliminary step for the creation of the conceptual frameworks and specification tables that are to be used as the basis for constructing the items. The IEA has developed different procedures to carry out such curricular analysis, both for countries whose national objectives are already explicitly stipulated, and for those where this is not the case. Part of the analysis is based on the information provided by teachers of a given stage or by studying school textbooks.
- Second, cooperative procedures are used in these studies in order to reflect existing cultural diversity. The most common practice in this respect is for national representatives to begin by proposing items in keeping with the agreed specifications table, which are then studied jointly to rule out any problems they may pose for particular countries. A subsequent pilot study is a key instrument

to check the items and to build the final test.

- Third, rigorous procedures have been developed for the direct and inverse translation of the items into the different languages of the participating countries, in order to ensure the levels of difficulty are maintained, over and above the specific language used for the test.

### *c) Need for Technical and Methodological Innovation*

The desire to explore levels of performance, differences between countries and national inequalities, coupled with the concern to explain the results, imposed a need to develop and use innovative statistical analysis methods. The adoption of LISREL-type analysis, and the Item Response Theory (IRT) approach, conducting hierarchical or multi-level statistical analysis and establishing *benchmarks* or international achievement scales, are just a few examples of methodological innovation apparent in IEA studies.

This innovation is evident not only in the methods and techniques, but it is also evident in the areas of evaluation, where an intention to study fields that have been little explored has become discernible. It must be acknowledged that most international performance evaluation has focused on academic areas that are considered fundamental and to which much attention has already been paid, such as language, mathematics or science. The IEA, on the other hand, has also sought, as far as possible, to broaden the scope of its research to other areas. Consequently, in its early years, the IEA included writing in its research subjects, although it did encounter major difficulties when it came to finding a standardised, reliable and inexpensive way of evaluating this. It did the same with classroom environment, which was studied in the early Eighties. Other areas explored with greater success have been early childhood attention and care, civic education or the educational application of information technologies. In all of these areas, IEA's contribution should be recognised and applauded.

## **IEA in the Nineties: A Period of Change**

At the beginning of the Nineties, when I joined IEA, major changes in education were taking place internationally, and these had a direct effect on the work of IEA and other organizations. The most noteworthy of these changes was the fact that assessing educational achievement became a matter of particular interest to many governments, which began to set up their own national evaluation systems during the Nineties. Early initiatives, such as those of the USA, Chile, France and Britain, were followed by those of Sweden, Spain, Argentina, Brazil and Italy, on varying scales and at different paces.

Moreover, countries began to compare their own situation with that of other countries or of their neighbours. The United States is a particularly revealing example, because its interest in evaluating its own performance in comparison to that of other countries was a contributory factor in promoting the initiatives set up by diverse international organisations. To put it bluntly, some of the outstanding IEA studies, namely TIMSS, would not have been possible without US political and economic backing.

What this process made clear was that the ideas that IEA researchers had put forward in previous decades were now relevant from a political perspective. The assessment of educational achievement was no longer simply a matter of concern to researchers, but also to those responsible for the education system, because the improvement of education performance and quality had become a burning political issue in an increasingly globalised world.

The concern shown by governments and education authorities had a number of significant consequences. The first and most obvious of these was that it imposed a need to design new systems to monitor the changes taking place in education, and which would also be capable of evaluating educational progress over a period of time. In addition to initiatives of this kind adopted by the major national evaluation sys-

tems, this demand also resulted in international bodies and associations being forced to act.

Thus it was that in the late Eighties, the OECD's Centre for Educational Research and Innovation (CERI) launched a project involving the construction of a system of international educational indicators. The project entitled INES was to define, compile and publish a set of indicators on the state of education in the organisation's member countries.

The indicators referred to different areas, covering the context, resources, processes and outcomes of educational activity. But despite its undeniable rigour, this conceptual approach did not solve all the problems. Indeed, it was outcome indicators that at first proved to be something of a stumbling block. Even though there was general agreement on the need to tackle this area, the fact was that the information available was very limited. Moreover, setting up a mechanism to generate indicators of this type was a laborious and complicated task.

On considering these problems, interest grew in the work the IEA had developed over several decades. On the one hand, it had collected data on the performance of education systems over many years. Although not all OECD countries had taken part regularly in this research, the list of participants was long enough to permit reasonable use to be made of the information. And although the purpose of the studies in question had not been exclusively, or even primarily, the calculation of indicators, their design definitely enabled these to be constructed without too much difficulty. IEA had also developed technical procedures that had been sufficiently verified to be used in international evaluation studies and which could provide the basis for the OECD to collect its own outcome indicators, should it decide to embark on such a task, as indeed it did at a later stage. Moreover, IEA had contributed to the creation of teams of evaluators in many countries, giving them a sound experience in the problems and practices involved in international evaluation studies. In short, it can be said

that the IEA contribution to this international, comparative task was by no means insignificant. Evidence of this is that the first publications in the *Education at a Glance / Regards sur l'éducation* series used outcome indicators taken mainly from IEA studies.

Those changes that had taken place over the years and the growing interest across the world in international comparisons also had other repercussions on IEA operations and projects. From the operational perspective, the most striking of these was the diversification in the profile of member states representatives. Whereas in its early days, national representatives had tended to be renowned researchers, some of these were gradually replaced either by the heads of new national evaluation bodies or the representatives of education authorities. This change was indicative of the interest governments were then showing in educational evaluation.

Another change that occurred on the Nineties was the growth in the number of countries taking part in IEA. The incorporation of new members from Europe, Asia, Africa and Latin America broadened the scope of the studies undertaken. The IEA's role as a forum at which political representatives, authorities and education researchers meet and work together, which it has always been anxious to preserve, was strengthened by having members from an ever-growing, more varied list of countries. Consequently, the association encouraged dialogue between regions with differing degrees of development, contributed to the transfer of accumulated experience and brought with it new learning for all concerned. Moreover, its openness towards less developed regions had the added effect of strengthening relations with entities such as the World Bank, which supported its research.

The impact of these new concerns also forced the IEA to be rigorous in its application of an idea which had already emerged in the late Eighties, namely the need to implement a cycle of studies. In other words, in contrast to individual studies not conducted at any set intervals, it had to design a series of studies that could provide perform-

ance assessment on a regular basis, at least for some of the core areas of the school curriculum. Indeed, the approval in 1990 of the regular implementation of the *Third International Mathematics and Science Study (TIMSS)* marked a major strategic turning point and represented the adoption of a way of working that still continues today. In my opinion, this change was a remarkable one, deeply affecting the IEA's line of work. The debates provoked during the design period of TIMSS, which I could observe in the Standing Committee meetings, were the most evident sign of this change of scope and time.

The construction of a cycle of studies enabled educational progress to be charted in a consistent manner over time. By repeating certain studies such as TIMSS (conducted in 1995, 1999, 2003 and 2007) or PIRLS (with data collection in 2001 and 2006) on a regular basis, progress data could be obtained which shed new light on the development of education systems. This contribution to the implementation of international education evaluation studies opened the way for other projects that are currently underway, including the PISA project.

### **My Time as Chair (1999-2004)**

At the Tel Aviv general assembly (1998), I was elected chair of the IEA, a role I effectively assumed at the end of the following assembly in Oslo (1999). Over four years, I was entrusted with a responsibility for which I feel immensely proud and in the course of which I was to learn many new things.

During this period, I was able to make the public presentations of TIMSS 1999, Civic Education 1999 and PIRLS 2001, debating these in a number of international fora. I also attended diverse meetings on behalf of the IEA and had the opportunity to take part in encounters with education researchers and leaders in different places around the globe. I had hands-on experience with the day-to-day running of international education systems evaluation, and worked with governments

in countries as diverse as Morocco, France and Argentina. It was, for me, a highly gratifying period, both from a personal and professional perspective.

My work was facilitated by a rock-solid structure and the support of many people I feel it is only fair to mention. In the late Eighties, the IEA's Secretariat was organised in Amsterdam, where it had a small team that was as competent as it was efficient. It was Tjeerd Plomp who started setting up this base - initially together with the first Executive Director, Bill Loxley, and as of 1997 with the highly efficient support of Hans Wagemaker. Hans's role as Executive Director of the organisation should be publicly acknowledged, because his sustained efforts provided the key to securing a good relationship with the government and financial organisations of the USA and with the World Bank. He was also able to set up an effective office staffed by the irreplaceable Barbara Malak. Of the people who have worked alongside her, Financial Manager Juriaan Hartenberg deserves a special mention. The above-mentioned and their successors have made possible the IEA's continued international presence and have ensured smooth relations between its members and other organisations. Without the Secretariat's work, it would not have been possible to keep up the pace of the IEA's development or the consolidation process that has taken place over the years.

A major development during this period was the strengthening of the Data Processing Centre, located in Hamburg since 1995 and which at this time underwent considerable growth. Particular importance must be attributed to the task undertaken by co-directors, Heiko Sibberns and Dirk Hastedt, and the rest of the Centre's staff. The DPC's growth and consolidation lay the foundations for the IEA's expansion and reinforcement, as well as safeguarding the quality of its studies, in addition to providing services to diverse bodies and projects. Although, at times, doubts were raised about the risks we ran by taking these decisions, time has proved us right in the end.



Also to be underscored is the important work carried out at the International Study Center (ISC) at Boston College, which has taken primary responsibility for TIMSS and PIRLS. Particularly worthy of mention is the quality and quantity of work undertaken by Al Beaton, Ina Mullis and Michael Martin, with support from other colleagues like Eugene González. Although I was acquainted with them before I became chair, it was during those years that I really had the chance to gauge just how vital and invaluable their contribution to the IEA's work was. The ISC has been the backbone of IEA studies since the Nineties.

I cannot fail to mention the directors of other studies who also contributed to boosting the IEA's international status. The work of Judith Torney-Purta and Rainer Lehman in leading the Civic Study was crucial, directing, as they did, a study that has had huge international repercussions and which has opened up a whole new field of highly topical and relevant study. The same must be said of the task completed by Hans Pelgrum and Tjeerd Plomp in the ICT in Education study, an area in which the IEA made one of the first international contributions. Although he was no longer IEA chair, Tjeerd continued to be involved in the activities of the organisation he did so much to promote.

None of this work would have been possible without the cooperation of diverse national representatives, solidly committed to the IEA, to which they gave their wholehearted support. Without seeking to be exhaustive and at the risk of leaving many out, I would particularly emphasise the constant presence of Christianne Brusselmans, always positive and close by, of Eugene Owen and Gordon Ambach, who did so much to ensure US support, of Constantinos Papanastasiou, always such a good friend to the IEA, David Nevo, Ryo Watanabe, Alain Michel and Gérard Bonnet, Leonor Cariola, Peter Vári and so many others. Their cooperation showed me that the IEA was not an organisation burdened by bureaucracy, but a true alliance among researchers, evaluators and heads of ministries, and one that could really help to improve education.

One of the most complex issues I had to face over these years concerned the relations between the IEA and the OECD. The implementation of the PISA project had significant repercussions for the work of the IEA. On the one hand, a certain overlap between PISA, TIMSS and PIRLS forced us to consider our future strategy. On the other, the work overload created for evaluation centres in quite a few countries by the accumulation of international projects, when added to the development of national evaluation plans, required us to consider the demands voiced by participants. Moreover, many national representatives were assigned to both the IEA and to PISA, which proved a double-edged sword. It was therefore an issue that took up a lot of time and attention.

I have to admit that, despite my previous experience working with the OECD and my proximity to many of its leading figures in the education field, relations were often far from simple. The IEA saw threats on the horizon and preferred to keep an open dialogue on future plans, which was not always easy or even possible. For its part, the OECD wanted to set the agenda in line with its own projects, which interfered with the work of the IEA. I do not wish to judge my own contribution to this task, because I do not feel it is my place to do so. I would, however, like to express my satisfaction that this complicated period now appears to have given way to a new phase, in which relations have normalised, where both organisations seem to have found their own space and where the IEA can continue to prove the value of its contributions to the evaluation of international educational achievement. In a field where cooperation should be the rule, it would seem only natural that cooperation agreements should be reached between international organisations working in the same area.

### **The Future of the IEA: Key Challenges**

In its first fifty years of existence, the IEA has firmly established its place in the field of educational achievement evaluation. If we glance

back and compare the current to the previous situation, the changes leap to the eye. The IEA has established itself internationally. It offers a relevant combination of regular study cycles—in particular TIMSS and PIRLS—with new examinations of less traditional areas, of which civic education is a current example. The IEA has become a referent for studies that evaluate educational achievement in connection with the school curriculum and with teaching and learning processes. It has extended its geographical scope to include regions of the world where it was previously barely represented. This does not mean, however, that it can afford to sit back on its laurels or that no challenges lie ahead.

In my view, indeed, the IEA has a number of challenges to face, and its response to these will determine its future development. First, although it would appear to have established study cycles, international education system evaluation is in a state of continual change. National evaluation systems have spread all over the world. They have grown and are now much stronger. Many countries now take part in current international studies, which have increased their scope, coverage and influence. But this expansion in evaluation activities has also aroused concerns. A substantial group of national education leaders now believe that education systems and schools run the risk of being over-evaluated and are calling for reasonable limits to be placed on the size of these activities. It is therefore highly likely that, over the next few years, the international agenda will have to introduce changes designed to address these new demands. We cannot, then, claim to have reached a phase of stability. Rather, it is very possible that we will see further change in the near future.

Second, although the IEA has made key contributions to the development of education evaluation studies, it cannot consider its mission to be complete. The development of new evaluation techniques and instruments is unstoppable, and both current and future studies will have to adapt to the new situation. Attention to the development of

skills and competencies and the need to assess them properly, for instance, should not be neglected in future projects. More specifically, the connection between school curricula and subjects and the competencies they help to develop should also be addressed.

Third, the IEA has always proposed going further when measuring achievement and probing for answers. Although this has been an ongoing objective, we have to admit that the results obtained have not lived up to the ambitions of their aspirations. It is true that today we know far more about education outcomes than we did fifty years ago, but we still have no empirical certainties about the factors that explain them. As I see it, this should be the key short-term objective, because our education systems need a set of rigorous guidelines that can direct our improvement processes.

Fourth, now that education system evaluation has undergone such expansion and diversified its actions, we need to turn towards the integration of evaluation systems in such a way that national and international studies interact and are complemented by other kinds of action geared to evaluating schools.

These challenges, in my view, are all related to putting into practice the idea that has always inspired the IEA and has become both its slogan and blueprint: the world as an educational laboratory. Although there can be no doubt as to its continued relevance, practical implementation has yet to be achieved, and it would be no bad thing if we could move in that direction in the next few years.



CHAPTER 25

**50 years of IEA: Some Reflections  
about its History**

*Judith Torney-Purta*

**Joining IEA**

My introduction to IEA was in 1967 at a conference at Lake Mohonk, New York at which comparative education specialists from England and the United States gathered to discuss the need for an expansion of the organization's studies beyond mathematics into six subject areas. I was a newly graduated Ph.D. from the University of Chicago who had recently co-authored a book based on a large survey of children's political attitudes in the United States (one of a handful of researchers who had published material relevant for the civic education subject area). When my professor/co-author, Robert Hess, couldn't attend the conference, I was invited. This was my first professional presentation, and I guess I did well, since I was invited to join the International Committee for Civic Education and a few years later ended up as the first author of the book reporting the study (Torney, Oppenheim & Farnen, 1975). It was during this study, in the early 1970s, that I spent a month in each of two summers working in the IEA offices in Stockholm. Soon after that John Schwille and the first class of Fellows supported by the Spencer Foundation arrived. They were absolutely critical to the success of the organization and its projects and to their long-term impact, but especially Jack has made contributions over several decades.

In those days, before desktop computers, to produce tables or figures required hand entry onto tables using results gathered from weighty computer outputs mailed to Sweden from Teachers College in New York. The civic education book itself was written on a typewriter during a dark December 1974 in Stockholm. John Schwille, Jo-Ann

Amadeo and I have reflected on the process of this first civic education study in the *Elsevier Encyclopedia of Education* (2010). Civic Education was the first subject area IEA studied where attitude measures were as important as knowledge measures. It was in some ways quite a risk for the IEA organization.

My involvement with IEA was then dormant until the late 1980s, when I was asked to serve on the Board on International and Comparative Studies in Education at the National Academy of Sciences in the U.S. (BICSE). This group, led by chairs such as Norman Bradburn, Emerson Elliott and Andrew Porter, was charged with advising U.S. Government agencies in order to regularize the U.S. participation in all international education studies (and sometime respond to their critics, see Bradburn, Haertel, Schwille & Torney-Purta, 1991). I remember when plans for TIMSS consisted of a few words scribbled in green ink on a transparency and depended on the perseverance of individuals like Tjeerd Plomp, Al Beaton and Eugene Owen. "The rest is history," as they say. I also remember when the decision was made to do case studies in the U.S., Germany and Japan and also the TIMSS-video study. These attempts to go behind and beyond the country rankings are an under-explored part of IEA studies. During this period I became interested in how all international studies were used by educators and policy makers and gave a speech entitled "Raising an Alarm: Is This the Only Role for Comparative Studies in the Debate on Educational Excellence?" (see Torney-Purta, 1990). John Schwille and I were also suggesting policy implications based on the civic education study of the 1970s during this period (Torney-Purta & Schwille, 1986).

In the mid-1990s I was asked by IEA to prepare a proposal for a second Civic Education Study, and had to resign from BICSE because I was now working on a study under their oversight. By that time TIMSS was thriving. As one BICSE member put it, "IEA went overnight from being a cottage industry to the Manhattan Project."

## **Key Persons of IEA at that Time, their Responsibilities, their Personal Philosophies and their Qualities**

My first IEA committee meeting was at the UNESCO Institute for Education in Hamburg in August of 1967. Neville Postlethwaite, the IEA Executive Director, soon moved to Stockholm as the organization found its new home there. Even in those early days what many refer to as “the Neville-working whirlwind” was in evidence. Lacking most of the technological resources we have today, with the use of typewriters, carbon copies and airmail, research coordinators got the material they needed from IEA in Stockholm. I always admired the straightforward empirical approach Neville took to research problems. If preliminary data showed that a scale wasn’t working, he’d call a group on the spot to sit down at a table to rewrite and simplify the items. I think it was a process like this that resulted in the scale of open classroom climate that has been a major predictor in the civic education studies.

Torsten Husen was the Chair of IEA during the period when I was first associated with the organization. He is remembered with respect and fondness, by both European and American educational researchers. He seemed to know what to say in difficult settings to establish rapport and cordiality among often contending parties. His offices at the University of Stockholm were IEA’s offices for a long time. At the same time he was devoting time to IEA, he was also becoming a “grand old man” among Swedish academics – a member, for example, of the Nobel Prize Committee. The early years of IEA were not easy, and both Torsten and Neville were always concerned about the organization’s survival, working diligently on academic and fundraising matters, and encouraging everyone else to do so as well.

Another major figure in statistical design in the early IEA was Gilbert Peaker, who would send long epistles to IEA Stockholm filled not only with statistical matrices and equations, but with quotations from Wordsworth and Shakespeare. I had the temerity of a young scholar to challenge Gilbert’s view that socioeconomic status of the home had to



be controlled (like handicapping a sailboat race, he argued) before the effects of learning conditions were examined.

Some of the world's most eminent scholars (Jim Coleman, Nate Gage, to name only two) passed through the IEA offices, each with his own valuable perspective on what comparative education research could provide to examine the then new comprehensive schooling movement and to stimulate educational improvement. In this period IEA was largely an organization of academics volunteering to participate in these studies because they were curious about the results and expected the knowledge gained to have important effects on education worldwide.

### **Responsibilities within IEA over the Past 50 Years**

I became the Chair of the International Steering Committee for the IEA Civic Education Study (CIVED) in 1993, after being invited to present a proposal to the General Assembly. I remember that the night before that presentation I was musing about how little was known about civic education in the post-Communist countries, which were trying to move toward more democratic education. I changed my prepared presentation to suggest a two-phased study, in which the first phase was a set of national case studies. This turned out to be a good idea, both because of the substantive issues that we learned about in different countries (allowing us to write better test and survey items) and because it allowed national research coordinators to have a chapter publication to show to their funders relatively early in the long process of conducting the study. The Pew Charitable Trust partially funded this phase (which included extensive collaborative item development), and this was the first involvement of Jo-Ann Amadeo, a colleague from Maryland, in IEA. We call it "the brick," since it is 627 pages long (Torney-Purta, Schwille & Amadeo, 1999). It won an award as a unique and high quality academic book from the American Library Association. An edited book of reflections across case studies was also published (Steiner-Khamsi, Torney-Purta & Schwille, 2002).

Founding members of the Steering Committee included, among other, John Schwille, Rainer Lehmann (whom I'd met when he worked on the Literacy Study) and Barbara Malak-Minkiewicz (whom I'd met at a civics curriculum meeting in Warsaw, who had never heard of IEA until she joined this committee). With funding from a German foundation, the International Coordinating Center was established at the Humboldt University of Berlin under Rainer Lehmann's leadership. Wolfram Schulz joined the project as associate coordinator a couple of years later. I won't describe the details of the findings but will say that we succeeded (in spite of relatively lower levels of funding than other projects) to meet IEA standards and to build a collaborative structure. My colleagues and graduate students at the University of Maryland's Department of Human Development helped us to meet our deadlines for the report on 14-year-olds in 28 countries (Torney-Purta, Lehmann, Oswald & Schulz, 2001) and on upper-secondary students in 15 countries (Amadeo, Torney-Purta, Lehmann, Husfeldt & Nikolova, 2002, both reports at <http://www.terpconnect.umd.edu~iea> ). Funding from the William T. Grant Foundation helped us during this period. We honored the academic roots of IEA by holding the report releases at professional meetings (in 2001 at the Comparative and International Education Society meetings in Washington and in 2002 at the International Society of Political Psychology in Berlin). Several publications summarized these results for different audiences (Torney-Purta, 2001; Torney-Purta, 2002, Torney-Purta, Barber & Richardson, 2004). Since those releases my graduate students have been publishing with me using techniques such as HLM to look at Latino students in the U.S. (Torney-Purta, Barber & Wilkenfeld, 2007), students' volunteering in 4 countries (Torney-Purta, Amadeo, & Richardson, 2007), support for human rights in 27 countries (Torney-Purta, Wilkenfeld & Barber, 2008), gender differences in efficacy and support for women's rights (Barber & Torney-Purta, 2009), and support for immigrants' rights in 25 countries (manuscript in progress). In addition, we began publishing short research summaries that can be found on the website

of the Center for Information on Civic Learning and Engagement (CIRCLE - <http://civicyouth.org>, including a study linking teachers with students found in Torney-Purta, Barber, and Richardson, 2005). Jo-Ann Amadeo and I also coordinated a workshop on data analysis using CIVED data for about a dozen researchers from low-income countries in conjunction with the International Psychology Congress in Berlin in 2008. The *Handbook of Research on Civic Education in Youth*, published by John Wiley in 2010 includes references to the CIVED study in three quarters of its twenty four chapters.

### **How IEA has Changed over the Past 50 Years**

The comment quoted earlier about going from a cottage industry to the Manhattan Project is actually quite apt. When I began with IEA forty-some years ago it was a loosely coupled network of academics who knew each other and each others' families. One of my early memories of IEA is in my small apartment's living room in the Wenner Gren Center in Stockholm in the early 1970s. We were having a singing party, with Kimmo Leimu and me trading Finnish songs and Ingemar Fagerlind helping with Swedish songs. My daughter Kathy was about 5 years old, and she fell asleep on Torsten Husen's lap. I remember this as part of the family atmosphere of IEA. I think this is less prominent today. Another change is that the center of gravity of studies has moved to research institutes, often those connected with ministries of education. Professors and doctoral students still participate, but it is my impression that this is more likely to be in secondary analysis than in the design and conduct of studies.

In the first international math study fifty years ago the aim was to explore the associations between various inputs to learning and outputs in the form of test scores. This aim was also central to other studies up through the mid-1990s (Torney-Purta, 1991). When U.S. Government agencies began to play an important role in funding studies like TIMSS, much of the interest shifted to country rankings. [Note:

this actually began with the *Nation at Risk Report*.] It was thought that policy makers would be responsive to the need for educational reform if they could see that the U.S. was falling behind other industrialized countries. Personally, I think that rankings are important but give only a partial picture. They get headlines, but give little guidance to educational practice and ignore many aspects of the context of education, which as a developmental psychologist I believe is critical. It is hardly the case that the U.S. could as a country adopt Singapore's math curriculum, for example, and transform student learning thereby.

We now have technological and methodological tools to explore the process of education not merely the ranked mean scores of countries. Given sufficient funding we could examine processes of education in multi-method studies, for example, with video projects in areas such as civic education. These could tell us a great deal that no test or survey can provide. I think it is important that areas other than science and mathematics remain in the mix of IEA studies (although I have resigned myself to the fact that they are unlikely to generate as much press or attention from governments).

### **The Contributions of IEA to the Field of Educational Research in the Past 50 Years**

IEA has demonstrated that it is possible to compare countries and educational processes within countries using outcomes of education (not just enrollment figures). IEA was the first group of researchers to use and measure Opportunity to Learn. The Technical Committee provided consultation to projects about modern measurement that has enriched educational research generally. However, some techniques such as IRT scaling are still difficult to explain to those in other disciplines who want to use CIVED data, particularly political scientists. In the last few years as getting permission to test students in schools has become more difficult using large scale data sets has been seen as having more value (even by psychologists who used to insist that students

collect their own dissertation data). IEA's policy of releasing data for use has always been a hallmark of the organization. In their training of researchers and in the networks established IEA has certainly broadened the scope of research relevant to civic education, and in other subject areas as well. The results of CIVED were released at about the same time that some political scientists and philosophers were becoming interested in civic education in the U.S. and Europe, and shaped that dialogue by providing evidence for promising practices in *The Civic Mission of School* (an influential policy paper), to give just one example.

### **Suggestions for the Improvement for IEA**

I think IEA could benefit from exploring several directions. First, it should pay more attention to its history, including some projects completed before TIMSS and PIRLS. An important aspect of its early history was the Spencer Foundation Pre- and Post-Doctoral Fellows. I think that a similar program (especially post-doctoral fellowships) should be explored for the current generation of individuals finishing their doctorates. Second, although governments are certainly an important audience for its findings, I think the balance has swung to cater to those audiences and away from professors, graduate students, and academic audiences in general. That balance could be reconsidered. Third, I think that carefully constructed measures of classroom environment (such as the CIVED Open Classroom Climate for Discussion and also perceptions of the Value of Participation at School) should be used in all studies. Fourth, I think that as much continuity as possible should exist between projects over time. That is, as few changes as possible should be made in measures that are working well. The old saying is appropriate, "if it isn't broken, don't try to fix it."

### **My most Memorable Moments in IEA**

I have both happy and sad memories. A happy memory was when I

was elected an Honorary Member in 2001. I was especially proud as the first woman to be so honored by IEA, and the award has a prominent place on my office wall. Also I was able to be at the 50th Anniversary of IEA, the celebration in Berlin and was on a panel with some people in the organization who had been valued colleagues and friends for a very long time. That brings me to the sad memory. I hadn't seen Neville Postlethwaite for about 10 years before that GA, and that was to be the last time, since he died a few months later. He was absolutely vital to the organization in its early years, and his powerful research skills were very important to its success throughout its history. I last saw Torsten Husen in 2007 thanks to a visit Ingrid Munck arranged while he was still in his Stockholm apartment crammed with old books and maps. John Schwille, Ingrid Munck and I organized a memorial at the CIES meetings in 2010 at which we honored the memory of both Neville and Torsten. Ingrid brought a touch of the old IEA spirit by providing a large bouquet of flowers and reading a moving Swedish poem.

### **My Activities and Publications:**

I have covered my responsibilities with the civic education projects of the early 1970s and starting in the 1990s above. Over the period from 1975 to the present I have co-authored or co-edited 6 books (reports of the civic education surveys findings of 1971, 1999 and 2000, two books about the case study phase 1 of CIVED, and a special report on the Latin American results published by OAS), and I also edited a special issue of a journal (Torney-Purta, 2007). I have also published 15 chapters, 3 entries in encyclopaedias, and 20 articles. In recent years most have been co-authored with my talented doctoral students. I have made more than 60 presentations about IEA (especially the civic studies) at professional conferences.

My IEA work was instrumental in my receiving the Decade of Behavior Research Award for policy related research supporting democracy (along with scholars in political science and communica-

tion and presented at a meeting held at the US Senate Building) and three years later the American Psychological Association's Award for Distinguished Contributions to the Advancement of Psychology Internationally. That APA speech (Torney-Purta, 2009) gave me an opportunity to present both a new way of making survey research more relevant to policy-makers (through cluster analysis conducted in 5 Western and 5 Eastern European countries with CIVED data) and to raise concern about lack of attention to the cultural and educational context in published research in psychology journals. I co-edited a Wiley handbook (Sherrod, Torney-Purta & Flanagan, 2010) in which the CIVED study receives attention in many of the twenty-four chapters. There is material cited relevant to cross-national views of political participation, to immigrants' education, to protest participation, to the role of classroom discussion, and to the use of multiple research methods.

In general my aims have been to help audiences in all subject areas to see beyond the country rankings to ways of understanding students' learning processes and their context and also to inform specific reforms in the area of civic education. Some specific policy-related efforts were these. First, after a meeting of Ministers of Education of the OAS in Trinidad about 4 years ago extensive initiatives were planned for civic education in Latin America and the Caribbean (based on a report based on the CIVED data, Torney-Purta & Amadeo, 2004). This initiative has resulted, among other things, in the participation of six Central and South American countries in the International Civic and Citizenship Education Study (ICCS09); with so many countries a special Latin American regional module could be justified and was developed. Both Chile and Colombia have undertaken specific civic education reforms based on findings from CIVED. Second, in the United States, I have written two papers for policy-makers, one issued by the Education Commission of the States (Torney-Purta & Vermeer Lopez, 2006) and one on 21st Century competencies issued by the American Bar Association in association with the Campaign for the Civic Mission of Schools (Torney-Purta & Wilkenfeld, 2009). The IEA

CIVED Study and its findings have contributed to a variety of initiatives decentralized at the state or local level in the United States. Third, the Council of Europe's initiative on civic education took account of many CIVED findings and the persons who formed its policy committee included several who had been active in the CIVED project.

### **In Conclusion**

IEA has been the source for some of my most valued friends whom I met decades ago through our IEA activities; some of my most creative ideas (deciding that we needed to have a case study phase for CIVED to find out what civic education meant, especially in the post-Communist countries); some of my most frustrating moments (which I tried to handle with determination tempered with grace); and, some of my greatest feelings of professional accomplishment (establishing a collaborative atmosphere among the participants in the CIVED project). I can't imagine what my professional life would have been like if I hadn't come to that first meeting in 1967.

### **References**

- Amadeo, J., Torney-Purta, J., Lehmann, R., Husfeldt, V., & Nikolova, R. (2002). *Civic knowledge and engagement among upper secondary students in sixteen countries*. Amsterdam: International Association for the Evaluation of Educational Achievement.
- Barber, C., & Torney-Purta, J. (2009). Gender differences in political efficacy and attitudes to women's rights influenced by national and school contexts: Analysis for the IEA Civic Education Study. In D. Baker and A. Wiseman (Eds.), *Gender equality and education from international and comparative perspectives. International Perspectives on Education and Society*, Vol 10. Bingley, U.K: JAI/Emerald Group Publishing. (pp. 357-394).
- Bradburn, N., Haertel, E., Schwille, J., & Torney-Purta, J. (1991). A



- rejoinder to "I never promised you first place." *Phi Delta Kappan*, 72 (10), 774-777.
- Sherrod, L., Torney-Purta, J., & Flanagan, C. (2010). *Handbook of research on civic engagement in youth*. Hoboken, NJ: John Wiley.
- Steiner-Khamsi, G., Torney-Purta, J. & Schwille, J. (Eds.). (2002). *New paradigms and recurring paradoxes in education for citizenship*. Amsterdam: Elsevier Science (JAI Press).
- Torney, J., Oppenheim, A. N., & Farnen, R. (1976). *Civic education in ten countries: An empirical study*. New York: Halsted Press of John Wiley.
- Torney-Purta, J. (1990). International comparative research in education: Its role in educational improvement in the U.S. *Educational Researcher*, 19, 32-35.
- Torney-Purta, J. (1991). Cross national research in social studies. In J. Shaver (Ed.), *Handbook of research on social studies teaching and learning* (pp. 591-601). New York: Macmillan.
- Torney-Purta, J. (2001). Civic knowledge, belief about democratic institutions, and civic engagement among 14-year-olds. *Prospects (A UNESCO Journal)*, 31, 279-292.
- Torney-Purta, J. (2002). The school's role in developing civic engagement: A study of adolescents in twenty-eight countries. *Applied Developmental Science*, 6, 202-211.
- Torney-Purta, J. (Ed.). (2007). Reflections on the IEA Civic Education Study (1995-2005). *Citizenship Teaching and Learning*, 3 (2), 1-82.
- Torney-Purta, J. (2009). International psychological research that matters for policy and practice. *American Psychologist*, 64, 822-837.
- Torney-Purta, J., & Amadeo, J. (2004). *Strengthening democracy in the Americas through civic education: An empirical analysis of the views of students and teachers*. Washington, D.C.: Organization of American States, (also in Spanish).
- Torney-Purta, J., Amadeo, J., & Richardson, W. (2007). Civic service among youth in Chile, Denmark, England and the United States: A

- psychological perspective. In M. Sherraden & A. McBride (Eds.). *Civic service worldwide: Impacts and inquiries* (pp. 95-132). Armonk, N.Y: M.E. Sharpe.
- Torney-Purta, J., Amadeo, J., & Schwille, J. (2010). IEA studying civic education. In P. Peterson, E. Baker, B. McGaw, (Eds.), *International encyclopedia of education (3rd edition)*, Vol 4. (pp 656-662). Oxford: Elsevier Limited.
- Torney-Purta, J., Barber, C., & Richardson, W. K. (2004). Trust in government related institutions and political engagement among adolescents in six countries. *Acta Politica*, 380-406.
- Torney-Purta, J., Barber, C., & Richardson, W. (2005). *How teachers' preparation relates to students' civic knowledge and engagement in the United States: Analysis from the IEA Civic Education Study*. College Park, MD: Center for Information and Research on Civic Learning and Engagement.
- Torney-Purta, J., Barber, C., & Wilkenfeld, B. (2007). Latino adolescents' civic development in the United States: Research results from the IEA Civic Education Study. *Journal of Youth and Adolescence*, 36, 111-125. [Winner of the Richard M. Wolf Memorial Award from the International Association for the Evaluation of Educational Achievement honoring the best article using IEA data published in 2007].
- Torney-Purta, J., Lehmann, R., Oswald, H., & Schulz, W. (2001). *Citizenship and education in twenty-eight countries: Civic knowledge and engagement at age 14*. Amsterdam: International Association for the Evaluation of Educational Achievement.
- Torney-Purta, J., & Schwille, J. (1986). Civic values learned in school: Policy and practice in industrialized countries. *Comparative Education Review*, 30, 30-49.
- Torney-Purta, J., Schwille, J., & Amadeo, J. (Eds.). (1999). *Civic education across countries: Twenty-four case studies from the IEA Civic Education Project*. Amsterdam: International Association for the Evaluation of Educational Achievement.

- Torney-Purta, J. & Vermeer Lopez, S. (2004, updated version 2006). *Developing citizenship competencies from kindergarten through Grade 12: A Background Paper for Policymakers and Educators*. Denver, CO: Education Commission of the States. <http://www.ecs.org>.
- Torney-Purta, J. & Wilkenfeld, B. (2009). *Paths to 21st Century competencies through civic education classrooms*. Chicago: American Bar Association, Division on Public Education and Campaign for the Civic Mission of Schools. Retrieved from <http://www.civicyouth.org>
- Torney-Purta, J., Wilkenfeld, B. & Barber, C. (2008). How adolescents in twenty-seven countries understand and support international human rights. *Journal of Social Issues, 4*, 857-88.

## CHAPTER 26

### 50 Years of IEA: A Personal account

*Kimmo Leimu*

#### **Beginnings and Early Influences**

I trained originally as a primary school teacher at the Teachers College of Helsinki, Finland (1959), and after spending some years teaching in a Helsinki secondary school, then in a reception institution for juvenile delinquents, and for several years in the Department of Psychology, University of Jyväskylä, I was appointed in 1967 at the Center for Educational Research in the same University as Research Assistant, employed by the Society for the Promotion of the Educational Sciences (Kasvatustieteiden Edistämisseura) to be the National Technical Officer (NTO) for Finland in the IEA Six Subject Survey (IEA/SSS), being responsible for the national research coordination. So upon entering IEA, I had ‘insider’ experience from working in schools for both normal and deviating youth, as well as more theoretical training and university-level teaching in my major academic field (psychology), together with education and special education. In my new appointment, I was responsible for (leading) the applied work necessary to adapt IEA instrumentation to Finnish circumstances, conducting the pre-testing, dry run and main testing phases, and preparing the data for dispatch to the IEA International Centre. This particular arrangement went on until end of 1969, and I continued from 1970 to March, 1972 as Researcher at the Institute for Educational Research, University of Jyväskylä (with duties as above). Subsequently I worked as Head of the School Test Bureau, where my duties included the development of instruments and promotion of the use of student assessment mainly in the field experiment schools. This was undertaken in preparation for the curricular decisions that were taken for

the emerging new comprehensive system of basic education in Finland.

### *Influential Personalities in the Early Days of IEA*

The key persons over time are no doubt better known through IEA documents, particularly as the authors of the main volumes of the IEA Six Subject Study (IEA/SSS). Personally I most vividly remember Torsten Husén, then Chairman of IEA, Benjamin Bloom, Neville Postlethwaite and Gilbert Peaker as the central figures, and as individuals with substantial and leadership authority. Their visions, ideas, solutions and general wisdom can be detected in much of IEA's work even today. The names of other influential persons can be read not only on the title pages of the main IEA/SSS publications. (e.g. Peaker 1975), but also in other inspiring volumes and articles (e.g. Bloom 1972, 1980, et al. 1956, 1971, or later by Husén & Postlethwaite, 1994). Their professional qualities are beyond my assessment, since they were leading figures on the wider international scene, while I was only a (keen) student, trying to absorb their conceptual thinking and modes of practical research operation. Such learning took place most naturally at both the Council Meetings (later called General Assemblies) and the National Technical Officer meetings (NTOs, later National Research Coordinators), where intensive conceptual-theoretical discussions and practical development work always took place. My most impressive influences were the wise worldview and constructive leadership of Torsten Husén (e.g., Husén, 1983; Husén & Kogan, 1984), Ben Bloom reputed for his 'Taxonomies' (1956), but also his humane educational thinking (e.g., Bloom et al., 1971; Bloom, 1972, 1980), the statistical mastery and clarity of Gilbert Peaker, and dynamic management by Neville Postlethwaite. Needless to say, each person involved in leadership roles in the Six Subject Survey and as authors of IEA publications represented persons of remarkable expertise and high esteem. Mere listing of them provides an overview of the impressive professional quality employed by IEA - yet most took on their work at

IEA on a voluntary basis and as side duties. This to me indicates the intrinsic interest and spontaneous enthusiasm evident in the pioneering spirit of IEA during those early years. The inspiring list, in addition to those mentioned above, included David Walker, John B. Carroll, Robert L. Thorndike, Arthur W. Foshay, Alan C. Purves, Gilbert de Landsheere, John P. Keeves, L. C. Comber, E. Glyn Lewis, Judith Torney, Bram Oppenheim, Carolyn Massad, Clare Burstall, Bruce Choppin, and Richard Wolf. All of them were world figure in their particular professional fields, but also in the fields of education, assessment, sociology, sampling, or statistics (witness, e.g., the work done in the Education in Measurement, Evaluation, Statistics, and Assessment [MESA] program at the University of Chicago, or the International Institute of Education [IIE] at the University of Stockholm). Quite importantly, these senior members had an extensive information network based on personal contacts, whereby pre-prints, rather than re-prints were exchanged among the devotees. This 'invisible college', as it was then called, may exist even today, in the form of pre-publications, or other informal exchange of ideas and information. But such exchanges were not only 'undercover'; one could learn a lot by listening to professional discussions going on at every meeting or within the group work sessions.

### *Reflections on the General IEA Philosophy and Mode of Operation*

The early years (1960s through the 1970s) meant pioneering work, where the 'founding fathers' (notably Husén, Bloom, Walker, Peaker, Thorndike, Postlethwaite, Purves, Torney) were discovering new ways to approach comparative education research. The 'roots' of their thinking may be detected in earlier generations of educators, such as William James or John Dewey, or the great Ralph Tyler; and we can see the influence of these IEA founders in today's evaluation work. Perhaps even more immediately influential were post-World War II trends in general societal ideals related to educational pursuits. The

'human capital approach' meant striving for greater opportunities and equity in institutional learning provisions for entire age groups. Among others, this involved avoiding (too) early decision points, which then would determine subsequent learning paths and eventually even career choices. This brought issues of societal and educational values and policies once again to the forefront, increasing demands for supporting educational and curriculum decision-making with solid empirical information, rather than philosophy, politics, or wishful thinking.

The spirit of the time also gave rise to fostering closer international cooperation, in order to attain better quality of life and learning. At the same time, greater mobility of both the labor force and university students brought practical problems in demonstrating comparability of examinations, vocational skills, and professional qualifications (see, e.g., Coombs, 1968). Particularly in the Nordic countries there was a keen interest in developing their education systems with comprehensive school ideals, thereby providing similar – and solid -- basic education for entire age groups. This meant venturing into unknown territory, and rigorous approaches and technical standards were needed to monitor the progress of these new developments. IEA ideas thus had quite fertile soil in which to grow new ideas. Because direct comparisons and simple country rankings were seen to represent superficial populist incentive values, more solid approaches were deemed necessary; IEA recognized the need to develop educational assessment with a research approach to face the challenges involved. These ideals were among the undercurrents that can be traced in the impetus to initiate and conducting IEA-type research work.

The viability of these early concerns and decisions was further supported by subsequent work organized by various other international organizations, notably the OECD, which first examined economic, societal and educational indicators, where aspects of student learning, especially outcomes, became (through 1990s) of more interest (Moskowitz & Stephens, 2004). There is no doubt that IEA set a pio-

neering example for studies related to student assessment, which under OECD/INES auspices grew into the present worldwide cycles of the PISA studies. (PISA meaning the OECD Programme for International Student Assessment.) Yet IEA has managed to sustain its role as an important research organization, which has its scientific role and ideals to cherish (see, e.g., Husén 1983; Husén & Kogan 1984; Purves & Levine 1975; Super 1967).

IEA studies typically placed the emphasis on rather practical curriculum specifications related to subject-matter content, the structure and substance of background questionnaires and methodological problems arising in the course of study. Areas that were not so keenly pursued were the various theories or social ideologies of welfare and development that exist within education, or society at large. Thus few alternative problems or theoretical approaches formed part of the discussions. These discussions were open, but were also necessarily limited to the views and decisions made in the early phases of IEA work, and most likely based on work already done in the leading countries, or on the particular interests of the dominating personalities. By the same token, one should remember that IEA represented a new empirical, quantitative and cooperative approach to comparative education, and due to the multitude of potential problem areas and the differing socio-cultural make-up of the participating countries, the mission was clear: the main focus would be on student learning, educational goals and curriculum content in particular subject matter areas, rather than socio-cultural phenomena. Thus not much time was spent seeking alternatives, or discussing a wide variety of competing approaches, particularly in terms of cultural or socio-political problems. Although there were some developments in this direction (e.g., Super, 1967; Plomp, 1992), they were few – this is evident in the relatively little time and energy devoted to understanding and accounting for particular national circumstances, i.e., as part of more general contexts with special histories and other characteristics of a broader ‘ecological’ framework.



In brief, while high technical standards have ensured the reliability of IEA work, the validity of findings in terms of societal processes are likely to have suffered due to the necessary compromises in the content. In IEA's defense, it may be said that – apart from the omnipresent time pressures – the studies and problems which at the international level are 'external' to particular countries in terms of planning, preparation and conduct, can nevertheless be interpreted and reported 'internally' by researchers at the national level. This has established an interactive working model for undertaking comparative studies with a cooperative approach. In addition, the modular approach (with national and international options) can improve the quality of original data in this respect (see, e.g., Leimu, 2004a).

### *Personal Experiences and Benefits in IEA*

Beginning with most tedious and 'down-to-earth' practical tasks as the Finnish NTO (National Technical Officer, later called NRC, National Research Coordinator), I was first involved in instrument translations, field testing, sampling and main testing duties. After some years became leader of the School Test Bureau, then the Evaluation Department of the Institute for Educational Research (University of Jyväskylä), and then from 1973-74 had the privilege of working as a Spencer Fellow at the University of Stockholm. This involved my rather intensive involvement in the IEA/SSS (Six Subject Study) materials and emerging publications, but allowed me to develop my own research ideas. My two years as a Unesco adviser in the supervision of education at the Ministry of Education of Lesotho, Africa, allowed me to apply both theoretical views and practical means, not only to the main task, but to its evaluation aspects in particular.

Returning to my 'home base' at the Institute for Educational Research in 1978, I was again back to multi-national evaluation work with IEA. The 1980s and 1990s were busy times, with several active research projects and other activities. We had the privilege of organizing the 1980 IEA General Assembly in Jyväskylä, Finland. It brought me spe-

cial pleasure to introduce Finnish expertise to IEA's international leadership, in this case the Written Composition Study. Our Institute even housed its International Study Center, with the International Project Coordinator (Sauli Takala) and other staff members coming from this center. For myself, these arrangements were supplemented by membership in the IEA Standing Committee (1983-87), nomination to chair the IEA Training Committee (1983-85), acting as a member of the IEA Policy and Procedures Committee (1983-85 and, the Publications and Editorial Committee (1986-96), and functioning as Editor of the IEA Newsletter (in 1985-1988). All this time, I was also a member of the National Subject Committee for every IEA project in which Finland was then involved. - In later years I have been able to apply the IEA approach as a model while doing studies in Botswana (with Torsten Husén in 1976), Zambia (in the 1980s), and Cambodia (in early 2000s). In all of these I owe much to my experience working with IEA .

In Finland, with only one or two persons initially devoted to IEA studies on a day-to-day basis, there was a danger of IEA work becoming too 'personalized', i.e., identified as 'territory' for those few who were doing the work. I soon saw the necessity of widening this circle of professionalism, so that more aspects of the complex cooperative work would have adequate coverage. It therefore became a deliberate effort on my part to involve others in the work, most of whom had specialized subject-matter training and previous experience in student assessment. It was a relief and blessing that a talented group of such researchers became involved over the years, and who then developed their expertise through active participation, eventually assuming further responsibilities in many admirable ways. They became active in new IEA (and later also OECD) projects, as well as National Assessment work, and progressed to the highest professional positions. It was therefore with satisfaction and confidence that I retired (in 1999) from duties and obligations involving both IEA and OECD/INES. By the same token, not only did technological-methodological experience and know-how develop among these Finnish IEA

researchers (witness, e.g., Pasi Reinikainen's dissertation that received the IEA Bruce Choppin award in 2009), but a good deal of serious 'evaluation thought' and widening of professional horizons also took place in connection with these and other studies. As a result, there has been an effort to change the prevailing *cult of assessment and grading* into a *culture of evaluation*.

### Developments in IEA over the Years

IEA started as a pioneering effort to explore the potential for undertaking rigorous empirical studies according to scholarly ideals and scientific standards. As part of this exploration, the role and effects of multi-level contextual characteristics were studied as important determinants of school learning. The conceptual model included societal, economic, educational and family circumstances. The consistency of findings is established by studying a variety of *national applications as replications*, which can be examined as such, or in light of comparative criteria, as represented by simultaneous findings in other national systems. I tend to view this multi-national, multi-level approach to understanding the relative importance of such antecedents in different societal environments as representing *basic research in education*.

The early years of IEA were times of limited resources, both nationally and to some extent internationally. Participation with representative samples of students in each target population varied from eight to 21 countries. The Six Subject Survey (IEA/SSS) was conducted with an international budget of one million dollars, which was considered a luxury in those days. Member countries were not necessarily equally well prepared in terms of investment in large-scale ventures, and (fortunately) IEA international did not charge membership fees. For this reason it was not always possible to join a project from the very beginning. For example, Finland could not join some of the repeat projects until the pre-testing phase, when it could be done mainly domestically and on a 'time sharing' basis, e.g., without travel costs. Today there

is a rather different picture, even if there is still not an abundance of money, or personnel in light of the challenges faced. . Overall, thanks to the work of IEA and OECD, educational evaluation at the international level has become an accepted part of our education system operation, and has thus gained approval also at tables where research plans are made or accepted and resources assigned. With the number of actively participating systems increasing up to some 70, recent studies and years have witnessed IEA's development into a more comprehensive organization both nationally and internationally, despite the fact that national fees have become an indispensable condition for participation. At the same time the projects have become more carefully controlled from beginning to end, with ever-increasing formal procedures detailed at every stage through planning, development, fieldwork implementation, and reporting. In recent years, an international quality monitoring element of test management has been added. An efficient data processing unit enables smooth state-of-the-art analyses of the massive data sets. All this has had the effect of costs rising to levels that could hardly be dreamed in the early days.

One of the most important program developments has been the establishment of *research cycles*, whereby some essential 'core' studies have been identified and their regular sequencing established as a long-term program. This arrangement offers opportunities for time-series studies at system level. It also provides *predictability*, which helps national centers establish their particular priorities and make long-term plans which usually include a variety of studies, both national and international.

This practice has both advantages and disadvantages, depending mainly on the frequency of such recurring efforts (cycle interval) and their harmonization with other research programs. Among these, and especially due to its 3-year cycles, the OECD/PISA presents the main rivalry to the IEA at the international level. Both of these impose serious challenges for funding and programming at the national level. Some of these are discussed below, under 'improvement needs'.

## IEA Contributions to the Field of Education

Issues of dissemination, impact, and effectiveness are an interesting, challenging and often overlooked aspect of research and its 'repercussions'. Much has been written about utilization and impact issues (e.g., Leimu, 1992; Weiss, 1972, 1979, 1980a, 1980b). Historical and policy considerations will necessarily come into the picture when defining the major research interests and issues, selecting latent and manifest variables for study, and later on, choosing and elaborating certain research problems and results for interpretation and discussion among the main audiences. Strategic approaches adopted in the dissemination and actual use of information in decision-making will also differ from one situation to another, depending on the variety of contextual elements and purposes. Since IEA deals with system-level phenomena, the major considerations will derive from the national policy and social contexts. This is important to understand, as it represents both opportunities and constraints for professional, political, as well as 'popular' interests. The chosen viewpoints reflect issues of the most burning concerns, and different focal points will necessarily arise in different countries and their special contexts. Clear demonstrations of such variety may be obtained from newspaper clippings referenced in the OECD Education Indicators - INES - work, where the focal points of interest chosen from the common database tend to be very different in each country. This demonstrates the variety of interests with which the same data may be examined by different audiences, as well as the potential 'idiosyncratic' effects which may result from the findings. It also makes it difficult to determine the particular impacts of any enterprise. Simple discussion is rarely possible. The benefits at the national level depend not only on the contents and results as such, but also on the coincidental relevance and timeliness of findings viewed in their particular 'historical' situation, policy concerns, on-going reforms and innovations and, thus, also prevailing public concerns.

Perhaps the major and *most lasting impacts of IEA* may be seen as relat-

ed to its main ideas and ideals: that it is legitimate and necessary for systems of education to follow the quality of their efforts, and that fair comparisons between systems are of global and national interest - and feasible. It is increasingly accepted that rigorous empirical information is the basis for doing the work, that student learning outcomes are the most important criteria, and that the study of contingencies underlying achievement in different socio-cultural environments constitutes the principal challenge. This however is not always sufficient to capture public attention or generate professional impact. For instance, in Finland it took several years before the policy, professional and public interests were duly awakened to international comparisons, to the extent that (loose) strategies were emerging for large-scale assessments. However, evaluation work at the national level has gained methods and momentum through IEA's example - which has developed the necessary know-how for still unknown future efforts.

It has to be recognized that these studies are conducted - and thus are intended to address - system-level phenomena, rather than producing much insight into instructional processes. (For example, there was IEA's Classroom Environment Study, which was cut short after less encouraging intermediate results.) The range of potential focuses is diverse, ranging from policy relevance to subject-matter expertise and psychological processes at the individual level, and from organizational and interpersonal skills to reasonable methodological understanding. Also, *any research study requires an environment which is favorable in terms of utilization context and expected impacts.* By definition, systems dealing with entire age groups are large and their steering is slow. In curriculum work and major reforms, decisions are made at several levels of system operation, and the time scale for significant changes may be defined in decades rather than years. The sheer amount of work implied by large-scale multi-national assessments is vast and time-consuming, and although it means dealing with carefully collected and analyzed empirical evidence, few direct recommendations can be suggested, and in any case the *actual benefits are not immediately visible.*

In this regard, the earliest IEA studies were somewhat 'premature' in Finland. The historical context of these studies did not favor active demand for the results, although some aspects, e.g., of the First Mathematics Study received public attention. But since the basic (policy) decisions concerning our forthcoming major educational reform had already been made, there was no turning back. As a consequence, there was little genuine interest in results obtained in terms of a system soon to become obsolete. It was difficult to seek or accept findings which would possibly give rise to other development efforts in this situation. Thus there were no immediate consequences. On the other hand, the benefits of solid research can be indirect, and sometimes unexpected, opening opportunities for users with different interests and purposes. It is for this reason that IEA data can retain their value and prove their importance decades later. Personally I tend to see those obsolete results as benchmarks against which subsequent IEA (and other) studies might well be assessed.

### *Some General Views Regarding Research Impacts*

*Educational contexts, and thus, reasons for assessment studies* are manifold; they are also potentially influential in considering the impacts of assessment studies. In Finland, several more general societal processes have had an influence on education, in terms of its basic premises and service offerings. The country is large, meaning that there are considerable regional differences in climate and access. There are large remote areas, but also dense urban environments. The main processes taking place since the 1950s have been a radical transition from rural to urban, from agrarian to industrial society, from poverty to welfare and affluence, and from basic production to planning, development and innovation activities. Trends in polarization of welfare to increasing equity followed suit, and today we are witnessing a gradual return to polarization. In terms of education, similar changes include curriculum processes developing from a top-down model to school-based decision-making and ownership, supported by core curriculum guide-

lines backed by governmental decisions. At the same time, developments in technology and communication have opened new avenues and opportunities for efficiency and ease of exchange and cooperation across the world. Educational developments have followed suit and led to some remarkable achievements.

These are by no means unusual developments, and they have had certain consequences for educational research as well, since all these changes have affected the way research objectives and content are targeted and the work itself is conducted. Thus, evaluation studies today reflect global trends in societal, technical and economic processes and changes. Thus also, the results of these studies contribute to documenting trends in world history, not only in reference to education, but to the human condition more generally. IEA-type comparative studies are based on the existence of variance among a multitude of general contextual influences. However, these may have rather different origins and status from country to country, and may not become fairly examined and interpreted in the studies. Among such general contextual features one may see the following:

1. *The historical context*: do research results manage to depict the reality in a valid and reliable way? Are the effects revealed by research great enough to be of practical significance? Are these findings timely enough to respond to the original demand and sufficiently relevant to be seriously considered in educational policy and decision-making?
2. *The decision situation in a change process* (see Stufflebeam 1976, 2004): (a) What is the amount of change to be accomplished, generally, and as a result of the evaluation research? (b) In terms of the information grasp, what is already known about ways of dealing with and controlling the decision situation?

The main analyses emerging from an evaluation study are related to exploring and establishing the effects of different contextual features on various criteria. The relevance of these determinants figures strong-



ly in judging the gains, i.e., the significance and impact of the research endeavour. Similarly, the judgement of the relevance of a study will vary depending upon the person judging it – i.e., the value is eventually in the eye of the beholder, although hopefully this is the active user of the information.

Below, several general aspects of context in education studies that affect the interpretation and use of results are briefly outlined:

- *Physical*: The natural, societal, local and personal environments, which provide certain concrete opportunities or constraints to student learning (such as climate, distances, access, or facilities, tools and learning equipment).
- *Socio-cultural*: Population density, ethnic, religious, or minority groups, various subcultures and their effects on homes, schools, teachers and children as students and citizens.
- *Political*: Philosophical or ideological system of values, which influences curricular goals and content; student selection processes, and resource allocation principles (notably, opportunities for learning, fair treatment and equity in terms of life chances).
- *Economic*: General national welfare and affluence, or distribution of wealth. These determine the human and material resources available to education. They will also affect curricular goals, instructional settings and various kinds of support for learning.
- *Technological*: Prevailing level of technology and the daily living and learning environment it enables. Quality of tools and level of skills needed for work or recreation
- *Human relations and communication*: Family and peer- group structures and processes, access to other people and sources of information, such as media.
- *Administrative*: System structure (levels of planning, implementing, and control), management and leadership, accountability and its control practices, power relationships and employment qualifica-

tions, e.g., the use of information in decision-making.

- *Pedagogical*: Curriculum goals and objectives, time allocation for studies and subjects, degree of freedom in making educational choices. Among educational resources, one may include instructional media, teacher education and qualifications, and as a result, teachers' professional skills.

Apart from demonstrating strengths and weaknesses in the quality of learning and the distribution of these assets among different social groups, the main analyses and interpretations emerging from a study are related to establishing proven contributions in terms of contextual effects of the above kind.

It remains for the national parties to determine the importance of established contingencies to local development efforts and decision making. Such contributions to knowledge-production are thus likely to determine the perceived importance of research. It is noteworthy that a good deal of this is in the hands of national authorities.

Another question is *whether good – or poor – research results are beneficial* for initiating development processes in the system. Again, as an example from Finland, while the historical situation was not favourable for responding to the (mediocre) IEA results of the early 1970s, in later years the excellent results in subsequent studies, notably PISA, have vastly enhanced the interest in and felt value of international comparisons, resulting in keen interest and hopes for better funding (at least for comparative research). At the same time the threshold of approval has risen, so that even slight deficiencies in quality are enough to raise public concern. On the other hand, poor results have never been hidden, because they can be presented as a basis for further needs.

Finally, from the comparative point of view, a problem remains *whether very similar or different systems should be compared* in absolute terms, or in light of elements in the structural models. On the other hand, this may bring in added variance, which is one of the benefits sought; on the other hand, this may lead to issues of fairness, relevance and cred-

ibility. These principles have consequences for the attention the studies receive and their impact. Any and all of these may influence both the needs and felt significance of evaluation studies for actual use.

### *IEA Contributions to Educational Research*

- *Rigorous sample-based comparisons* and sophisticated statistical methods have become the hallmark of state-of-the-art, system-level educational research. IEA can be considered a paragon in this field. Not only policy makers and ministry officials, but school heads, teachers, and even students have gradually developed a sense for quality work in the area, and require a good rationale, impeccable instruments, and well-disciplined conduct of fieldwork for their efforts. Timely and useful feedback of results is also expected at every level. All this has established new standards, which have been set forth and developed in IEA and PISA work. As a forerunner, the main credit for such developments is due to IEA, which has made important contributions in the field.

In the case of Finland (joining IEA early on), much of the original drive was based on the insight of a few key persons, such as professors Martti Takala, Juhani Karvonen and Pentti Pitkänen (see Karvonen 1974, Kuusinen, 1967; M. Takala, 1963). Concerning the potentials of these studies – in addition to scientific curiosity – the main expectations for gains were related to *rather technical benefits*: How to conceive, prepare, manage and analyze results of large-scale evaluation studies. This methodological ‘know-how’ viewpoint was natural with the advent of our major school and curriculum reforms (in the 1960s and 70s), where major research efforts connected with field experiments were foreseen as necessary elements – and thus challenges to be faced. Because specific national contexts and policies were deemed fundamental to the eventual usefulness of any studies and findings, the main sight in my understanding was initially set more at *future national studies* rather than sustained participation in multinational research projects. In other words, apart from the problems inherent in comparability of educational outcomes, relatively little was expected to be

gained from substantial or theoretical findings other than certain practical and technical means – i.e., overall methodology. Thus initial expectations did not come with too high comparative-theoretical ambitions, but considering the situation and resources available, they were realistic. One may well say that much more than these original expectations were gained,, although competency took perhaps more time than was originally expected. More ambitious aspirations arose later, with the proven success of IEA studies and growing interest in international research cooperation in general. The work was indeed pioneering,, especially considering the somewhat indifferent early attitudes of headmasters, teachers, and even policy-makers.

At this ‘honeymoon’ stage, money was always tight, so it was not easy to secure resources for all requisite IEA activities. Apart from funding received through the research budgets of the University of Jyväskylä and the National Board of Education (NBE) (including assessments other than IEA), traveling expenses were separately paid by the Academy of Finland. Yet participation in IEA-type active working meetings tended to lose out against the more established academic conference trips, where papers are read to coincidental audiences, often without any further responsibility for follow-up research or development work. . For a relatively long time, the daily ‘grass-roots’ IEA work was done by one or two persons, and the ‘soil’ in which it was embedded was thin. Although this situation gradually improved toward the 1990s, it took the example of OECD/INES to show the way -- through intergovernmental agreements -- for providing adequate and stable resources for organized international work in educational evaluation. However, IEA projects have managed to maintain their place on the national assessment agenda, even if they are not based on such government-level agreements.

The Finnish IEA experience has been rather thoroughly covered and illustrated in a publication edited by Leimu (2004b), in which some general aspects of evaluation and international assessments are also discussed and chapters on a variety of substantive areas covered by

IEA are produced by several other authors. The book was dedicated to professor Martti Takala, the Founding Father of IEA in Finland.

- As a result of IEA's research approach, whose feasibility and usefulness has been proven in practice, a *pool of know-how and skilled evaluators has developed around the world*, all weathered in international and national research and development ventures. Such experience has provided insight and practical skills for conceptualizing and conducting demanding pieces of research with a quantitative empirical approach. Apart from the IEA studies, and perhaps equally important, is the fact that many researchers have developed the research skills and know-how through their active involvement in IEA, and these have been implemented in a variety of *non-IEA studies* conducted around the world. Such *secondary effects of IEA work* represent seminal capacities for further work, allowing different research tasks to be tackled and new applications to emerge in a variety of environments. (Examples of these include the SACMEQ studies in 15 African countries, led by Ken Ross and Neville Postlethwaite, those by Postlethwaite in Vietnam, and several others.) Activities like this are important spin-offs of IEA, because they are likely to be part of national development efforts and thus can significantly impact social progress.

- In addition to research competencies developed in the course of IEA's large-scale studies, one must recognize the *significant contributions brought about by the rich* databases compiled within each project. These pools of information, organized and openly available, can be considered *real landmarks as historical documents* of the state of education from early 1960s to this day. They contain multi-level and multi-faceted data and are a unique source of evidence, the like of which cannot be easily or reliably recovered from any other source. Such data becomes suitable material for method and theory development, as demonstrated e.g. by the work of Richard Noonan and Herman Wold (1976 and 1983 respectively). Initiatives such as the Bruce H. Choppin Memorial Award may be regarded as fitting spawning ground inviting continuous developments in the field of methodology. As already mentioned,

even if the timing of fieldwork has not always served current needs and interests, immediate impacts are not the sole purpose or benefit of carefully collected data. I am thinking particularly of the Finnish situation during the Six Subject Survey, where IEA data captured snapshots from the old school system. The 'tsunami' of the new basic education system soon wiped out such possibilities ( seeing 'the way it was'). At least some empirical light can still be shed on a variety of issues, whenever claims are made of 'the good (or bad) old days' as the case may be. Without this 'obsolete' data such possibilities would be definitely out of reach.

- I am tempted to mention one particular instance where IEA missed a valuable early opportunity for further pioneering and wider public interest (perhaps also better resources). This was when an initiative proposed at the Paris GA in 1979, which suggested that the collection and reporting of IEA data should be developed with an *educational indicator orientation*, was rejected. While this proposal was not supported, it was later taken up and further developed by the OECD as an element of its INES project (Indicators for National Education Systems) and its PISA extension (Programme for International Student Assessment). The latter has become a success story and its reports are bestsellers among OECD reports. - It is only fair to note, however, that IEA work in student assessment served as a model and inspiration for PISA in its early stages, although the OECD has since made impressive steps in developing its assessments and reporting in other directions as well.

At the same time, it is important to note that indicators as such may remain superficial if no in-depth analyses or explanatory discussions are involved. Thus the challenge remains for IEA to retain its role as a *bona fide research organization* by taking care of such activities - which it has done. But considerable attention to theoretical orientation, sufficient opportunities for in-depth use of research data, and renewals other than statistics might well be necessary for IEA to maintain its birthright in international research endeavors.

### *Some Further Research Considerations*

- *PISA vs. IEA*: It is clear that OECD owes an intellectual debt to IEA, even if the former has become more visible in the media and public discussion. PISA developed as a necessary consequence of the OECD/INES (Indicators for National Education Systems) Network A need to produce indicators of student learning outcomes. The original intention was to make use of IEA results (which did in fact occur in the early phases of INES). This might have secured IEA a stable role as a data supplier in indicator work. However, the frequency and timing, and partly the content orientation of the two parties did not match, so OECD decided to establish its own program: PISA (Programme for International Student Assessment). Its research cycle model (already established in IEA) has more frequent periodicity than IEA, but its contextual content is less comprehensive. Thus the IEA model provides possibilities for a fuller model of school learning. IEA would be wise to make full use of this situation by staking its claims as a research organization.<sup>1</sup> It could well spend energy doing this, rather than adopting overly tight cycles of fieldwork. This is because significant educational changes rarely take place within a span of few years. Systems and programs therefore need a fair chance to prove their viability in terms of practical implementation, and ample time for actual results to emerge. While serious curriculum changes tend to require some 10 years to become duly implemented, it may be a waste of resources to secure data for (premature) intermediate stages. The problem remains how to gear assessment cycles to the actual situation in the participating systems, especially as they do not march lock-step in their national development efforts.

- *National and international policy studies harmonized?* What began as a tentative exploration of the possibility of a unified empirical approach

---

1. Note, however, that more extensive context data is available through other OECD's sources, due to the scope of its versatile economic and social information gathering programs other than INES or PISA.

to comparative education has developed into more established forms of evaluation research cooperation. These developments have occurred at both the national and as the international level, while system-level assessments have become an accepted element of responsible system operation. What has been said above about the opportunities and constraints set by historical and other contextual contingencies will apply here. There may be further ways to develop the sensibility of large-scale assessments, one of them being the *distribution of research focus between national and international research actors*. One can hardly expect international studies to completely exhaust relevant problem issues. Often more questions are raised than answered by the big studies. However, such 'unexplained' elements could well be investigated through national efforts focusing on specific problematic points that need deeper, more intensive attention. This issue of *strategic approach* has been discussed by Bryk and Hermanson (1994) under the concept 'information pyramid' and further developed by Leimu (2001a, 2001b). In essence, this approach suggests that problems/issues detected in international studies be pursued in further depth at the national level, where the contextual and comparative findings of the international study would inform their research design but where local, supplementary information would also feature. Such potential of study are further discussed in a publication on a joint seminar held in Jyväskylä in 1998, which involved researchers, decision-makers and policy representatives from both IEA and OECD/INES, as well as Finland and Sweden (see Leimu, Linnakylä & Välijärvi, 2001). This stepwise and logical linking of national and international studies would conceivably not only 'cross-fertilize' research thinking by suggesting agendas for national researchers or students, but perhaps represent a further step in improving the effectiveness and usefulness of research.

- A well-considered *conceptual framework based on latent variables* is indispensable for serious research purposes, and IEA cooperation brought such needs and practices to the forefront. In so doing, it set an



example to national-level efforts, and also generated critical thinking on certain related problems. Nice examples are found in the works of Keeves (1972, 1995) and Noonan (1976, 1983). Whenever concrete definitions of underlying (latent) concepts are involved, various *approximations and compromises come into the picture* at the manifest variables level. At this stage, important national aspects may be compromised or ignored unless the original notions are well understood (socio-economic, or socio-cultural status constructs being a case in point). Thus, variables which are considered important in one system may receive less attention or approval in others. Unless this issue is resolved, eventual interpretations may lack equivalence and lose credibility, or be misleading. Such dilemmas are sometimes due to different values or ethical imperatives in the participating countries, their differing societal realities, language connotations, manifest images of the content, or simply differing personal opinions among researchers and their hidden agendas (e.g., 'favourite sons'). Depending on the sometimes delicate decisions taken to ensure the necessary compromises, the final choice of variables may therefore have different validity in different countries, due to their manifestation and thus (real or perceived) relevance. In short: unless differences hidden by consensus decisions are carefully considered, the particular variables included in the models and instruments may reflect national situations differently and may therefore affect the results and/or conclusions. The situation may be relatively simple in the most general system-level analyses (such as country rankings), but become more delicate when process variables such as educational selection, specialization, or qualifications are formally modelled, or when dealing with 'lower-level' concepts in the usual theoretical perspective.

- *The International Encyclopedia of Education*, a massive 12-volume general publication with several specialized sub-domains, including dozens of education system descriptions, represents a major achievement by presenting, among other, scholarly articles on important conceptual, technical and substantial issues in education, as well as edu-

cational research and evaluation. Its main editors were two leading IEA individuals: Torsten Husén and T. Neville Postlethwaite, while other prominent members of the IEA family were well represented in the initiation and as authors of the numerous volumes and chapters of *The Encyclopedia* (Postlethwaite & Husén, 1994). This remarkable achievement is a landmark of joint international efforts and may be considered a response to the need for a general and theoretical discussion mentioned earlier in these memoirs. In terms of research impacts, these volumes make an enormous contribution to the educational research literature.

### *Suggested Needs for Improvement for IEA*

As much as there is to commend the original ideas and work accomplished over decades, there is always room for improvement. A listing of some pertinent points is available in John Keeves's 1995 publication, "The World of School Learning". While endorsing practically every point brought up on p. 60 of this report, I have made an attempt to provide some further arguments for the educational needs listed by Keeves.

- The need for a *more comprehensive social-theoretical* grasp of the education situation in member countries, in terms of concepts and contents that are meaningful to the aims of comparative education. There is perhaps something to learn from the early comparativists' ideas.
- The need for a deepened *understanding and clearer picture of national circumstances*, including historical developments, so that the descriptive data can more accurately reflect the 'larger picture' and the data analysis can more authentically reflect the cultural, historical and ideological reality. This would necessarily mean a deeper involvement of local expertise in the process, and although it might cause practical difficulties, it would improve both the validity and credibility of findings and prepare the ground for future research ventures.

- The need to counterbalance the various *compromises* that become necessary if complete consensus is pursued regarding research issues and instrumentation. If some issues or variables are regarded as 'impossible' or the 'results [deemed] not publishable' in some countries (as has been the case), and thus some potentially important aspects in the greater picture are omitted, this may negatively impact on a truly comprehensive conceptual model. If international core variables in the questionnaires are to be kept to a minimum (as in indicator approaches), more space should be allocated to international and national options, which would retain better possibilities for more valid studies.
- The need for *regular study cycles*: While it may be advantageous to pursue very frequent research cycles, certain difficulties inherent in this can actually be disadvantages. Although the principles have been developed and agreed at the international level, problems tend to arise at the national level, where frequently occurring research obligations clash and may become an unreasonable burden for participating schools, teachers and students. These demands place similar strains on funding options. We have seen already how the congestion of studies can diminish the willingness of schools and teachers to participate in further fieldwork--this situation has been observed, at least in small countries, where the same schools are 'harnessed' to the tasks over and over again. This situation may eventually endanger practical possibilities for conducting large scale studies not only by IEA, but by any of its concomitants.

An equally - if not more - harmful consequence is the situation in which national centers and their researchers are kept occupied with ever-new data collection cycles, rather than allowed the time to make more in-depth analyses and reports, which are the true marks of scientific work. There is rarely the luxury of employing a second set of researchers to substitute those who are already working on a research project; this first 'set' must take up the second project without contin-

uing with in-depth analyses and publications. Yet significant research findings are the “raison d’être” of our studies, and I dare suggest that this may soon develop into a life-and-death struggle for (too) frequently occurring research obligations, and their eventual contributions to the world of school learning.

This should be noted by the IEA, if it aims to maintain its role as a genuine research organization. Large, international volumes are not enough to keep up its reputation; fewer reports, perhaps, but of more depth and scope, are what is needed.

In my humble opinion there is reason to plan research cycles in terms of the basic nature of system-level educational processes. The ‘ship of education’ is big and turns slowly, so it is a waste of effort and resources to adopt an *unreasonably rapid succession of international studies*. Major curriculum reforms and developments (which are likely to be of high public interest and imply further decision needs) may well benefit from assessments conducted at, say at least 5-year intervals. Less central domains and problem areas could well do with less frequent efforts. In so saying, I am thinking of the variety of *national assessments* that are scheduled and that can offer supplementary or intermediate data.

By the same token, it should be noted that reconciling national needs with the large-scale and curriculum assessments aims of international studies must be deliberately and carefully planned. Discrepancies/overlap in timing and content/focus may well make it difficult to attain significant results in this regard.

- Fair comparisons of countries could well inspire *comparative in-depth analyses* among systems whose contextual environment is rather similar, versus those that are very different. Although availability of greater variance is a viable aspiration, its benefits should derive from a truly scholarly effort to understand educational phenomena. Comparisons among systems with very different (economic, social, cultural, etc.) contexts are likely to cause misgivings among those less

privileged, and may sometimes even offer a misleading picture of the contingencies, i.e., the explanatory elements. Shortcomings should not necessarily be highlighted as such, but rather used in the explanatory models. In so doing, the question arises whether any explanatory analyses between systems have been done where the desired 'wider international variance' has been studied in terms of a single overarching model and data. Analyses in terms of national replications would seem to miss this opportunity.

### **About the Future of IEA**

- There is a degree of *competition among the major international actors* on the scene (IEA, OECD/PISA, and others), but there are also challenges vis a vis national assessments. It is obvious that IEA has been a model and inspiration for most of these, reflecting the currently fashionable drive for continuous development, or effectiveness and efficiency, for better or worse (see, e.g., Owen, Stephens, Moskowitz & Gil, 2004.) One might wish that IEA would find and retain its particular role and modus operandi as a scholarly/academic organization, and develop its work accordingly. One problem in accomplishing this is being creative and finding novel research approaches, while seeking stability in repetitive studies and longitudinal trends over time. Bridge items and calibration procedures may provide a solution, but raise questions as to whether more fundamental changes in strategy would be possible: Opening new target areas? Finding new strategic moments for life-long education research? Supplementing 'hard' data with 'soft' interpretative observations? Financial constraints no doubt will be quite significant here.

- It should not be only *international assessments* that are taken into account when conceiving the future of large-scale assessments. Because of policy makers' enhanced professionalism, teacher education and research know-how, *national governments and agencies* have followed suit and developed their own evaluation systems. These

have their advantages in being more specific and timely, and perhaps more focused on current discourse in particular problem areas. In Finland their subject-specific content (especially in languages) is particularly appealing to teachers. Having less rigorous obligations for quality control, but being 'domestic' and cheaper, they are considered more as tools, and can compete with international surveys, even if these have their international assets and thus more prestige in the eyes of the general public. From the students' and teachers' point of view, they are likely to diminish the 'aura' of international studies and constitute a challenge in terms of resource allocation, including school and classroom time.

- There seems to be a need for some *harmonization in terms of frequency* and content in the cycles of international surveys. This refers particularly to the situation with the two major actors on the scene – IEA and PISA. Although these follow somewhat different approaches in their instrument specifications and schedule, there is a good deal of *overlap among the core content/substance*, resulting in a sense of repetition. At worst there may be field stages where both studies are testing the same substantive content at the same time. As a result, there is over-burdening of some of the vital elements in the research program for schools, teachers and students. The current good will and patience regarding these obligations – seen as 'extra duties' – is gradually diminishing. With growing 'testing fatigue', obtaining co-operation is becoming increasingly difficult, as witnessed by growing rates of refusal for research requests.

### **Some Memorable Experiences with IEA**

- *International Curriculum seminar at Gränna, Sweden, in 1971.* During this six-week seminar teams from 23 countries were represented, involving special expertise in four-five curriculum areas, one of them being evaluation. World-class leadership and expertise were represented in the staff, with Ralph Tyler, Ben Bloom, Torsten Husén and

Bom Mo Chung as the main plenary speakers. Several other experts of high caliber (Havighurst, Blaug, etc.) also made contributions, leaving a lasting impression on the entire audience. In addition, smaller seminar sessions were part of the daily program. I felt it a special privilege to be a member of the Evaluation group, under Ben Bloom's leadership. Having been yearning for theoretical ideas and discussions in the field of evaluation, the Gränna experience offered me a great deal of intellectual nourishment.

- *NTO/NRC meetings*: These were always active working meetings, in contrast to most ordinary conferences, where papers are read out for an occasional (although no doubt interested) audience. As it was not always possible to convince national officials familiar with the more traditional concepts of international cooperation to provide funding for IEA-type working meetings, some of these had to be missed. At the meetings I attended, however, there prevailed such genuine enthusiasm that a keen working atmosphere was experienced at practically every session, with real problems and issues to discuss, experiences to share, and ideas to develop. All this was enough to create congenial feelings among the participants, resulting in true friendships across borders. Truly an 'IEA Family' atmosphere! - One cannot deny the excitement that was always associated with IEA meetings held in different countries and continents. The hosts were always generous in offering glimpses into some of the most interesting aspects of their culture - thus providing memorable lessons in international understanding!

- *General Assemblies* had the same active working spirit, where the senior members had an important leadership and advisory role. Regarding understanding of the basic rationales and consideration of various research ideals or strategic issues and ways to proceed, these were always valuable opportunities for learning and sharing. In comparison to the NTO/NRC meetings, and covering several on-going projects, the discussions necessarily took place at a more general level. This situation was thus educative in a slightly different way. Yet I will never forget the enthusiasm with which, for example, the Singapore

GA participants swarmed around a portable blackboard, some kneeling down to write their ideas, others reaching over them to eagerly provide their modeling suggestions. Truly an open invitation for all to contribute! I also learned a lot through my own membership in the IEA Standing Committee, where important procedural and policy issues were discussed.

- *Spencer fellowships*, which were granted annually for several years in the mid-1970s for two senior (post-doctorate) and two junior (pre-doctorate) candidates, provided opportunities for researchers to concentrate on their academic work under favorable working conditions. The seat was the University of Stockholm. Since the timing of my own fellowship year (1973-74) coincided with on-going work with the IEA Six Subject (SSS) volumes and IEA data bank, part of the time was devoted to joint efforts in checking and improving the quality of IEA/SSS international data and reporting. The role of other Spencer Fellows was also memorable in terms of educative academic discussions and the development of personal friendships. The Spencer experience allowed me to personally develop curriculum- and OTL-related (Opportunity-to-learn) ideas, which has meant a lot for my conceptual thinking ever since. Subsequent work obligations eventually prevented a full-fledged articulation of these ideas in research practice.

- *East-West Center seminar* in Honolulu, in January 1981, represented an effort by an outside institute to learn from the experience of a large-scale international venture. Articles were prepared by active seminar participants, as well as others experienced in IEA work. During the seminar, long daily sessions were conducted on a number of key issues related to conducting a multi-national project (IEA), supplemented by intensive interviews by East-West Center staff. The working atmosphere was diligent and productive, the resources and future plans looked promising. However, within a few months following the seminar, the East-West Center was dissolved and the work disappeared into oblivion. Recent inquiries directed to the recently revived East-West Center hoping to trace the outcomes of this work have failed, so



that obviously none of the interviews and only some of the contributing papers remain available. However, that experience was one of the highlights of my IEA-related work.

On a more personal note, might add that on my return from the Honolulu seminar I made a stop-over to meet a colleague and some friends in Washington, D.C. The visit coincided with the inauguration of president Ronald Reagan, which I was able to see and hear live, although from a distance, before catching my train to NYC from the nearby Union Station. The experience was remarkable in the sense that the closure of the East-West Center was to be one of the earliest decisions made by president Reagan. Little did I know that at that very moment, it was the beginning of the end for one piece of important IEA work.

- Sustained collaborative work, even across borders, builds up not only professional, but also *personal relationships*. At the regular IEA meetings (both General Assemblies and NTO meetings) *colleagues became friends, and long-term friendships were formed*. By virtue of my early involvement, I had the unforgettable privilege of meeting most of the founding fathers, and even being accepted as a colleague and co-worker. After giving due recognition to these pioneers and their ground-breaking ideas, one can identify the first group of NTOs/NRCs as representatives of the *first generation of full-time workers in IEA*. I cannot deny feelings of pride for belonging to this group, who had to struggle through mostly new ground and do so with relatively primitive tools. A good deal of learning took place at those gatherings, as a number of future leaders and top-grade researchers were there to share their wisdom and guidance. Among these, I might mention Malcolm Rosier, Richard Noonan, Gerry Pollock, Roy Phillipps, Zoltán Báthory, Clare Burstall and Wendy Keyes. Needless to say, close friendships were developed among *national IEA colleagues* as well, of whom I mention with pleasure and gratitude, Hannu Saari, Erkki Kangasniemi, Pirjo Linnakylä, Sauli Takala, Anneli Vähäpassi and the 'invisible' data manager Anja Poutiainen. There are of course countless others who deserve mention.

*The spirit of sharing* - not only of problems, but also ways and means - has helped to develop true respect and friendships among all those involved, at home and abroad. Underlying much of this feeling is the recognition and appreciation of the cooperative work effort that has been part of the *IEA family* spirit for as long as I can remember. What remains topmost in my mind is the conviction: *This kind of work cannot possibly be done alone!* This sense of a common mission presumes genuine respect for the contributions of every member in the entire enterprise: *Professionally and personally, it is what and how you do it, rather than who you are, that counts.* The feeling is not only affective - it is a 'conditio sine qua non' for the entire effort - the real IEA lifeline without which success would have been impossible!

On the other hand, there are also the more personal memories. Proofs of the depth of friendships are the written communications (often informal messages, or in terms of pre- or re-prints), greetings, and later on, *visits to and by former IEA colleagues*, many of those dating back 40 years. All of these reunions have been happy moments indeed (with Zoltan Báthory, Malcolm Rosier, Cristina Rodriguez, Richard Noonan, Roy Phillipps, etc.). Inevitably, these have lately meant moments of sadness and longing, due to the passing away of several good friends (thinking back to, e.g. Neville Postlethwaite, Roy Phillipps, Robert Liljefors, Torsten Husén, Alan Purves, and others).

- Last but not least among the memorable highlights: *Honorary membership in IEA*, granted to me at the General Assembly in 2009. This was indeed an unexpected and most pleasant distinction. It made me proud, thankful and humble at the same time, and admittedly, quite emotional. Definitely a final highlight with IEA.

## Final Remarks

I proposed at the outset to *give recognition to IEA-type work beyond and outside the actual IEA projects/studies*. A vast amount of work has been done using the IEA approach and conceptual frames *in different studies*

*and in different environments and throughout the world.* These activities may in some cases equal the amount of work done under IEA auspices proper. I am thinking, for example, of the work undertaken by Neville Postlethwaite and Ken Ross in 15 sub-Saharan countries in Africa, work done in Asia, and work carried out in the French-speaking countries of Africa by (French) people also involved with IEA. I myself have had the honor and pleasure of such experiences in preparing research plans and instrumentation for follow-up and impact projects in Botswana (under Torsten Husén), inspectorate work structure in Lesotho, Zambia's curriculum and learning materials project, and most recently in Cambodia with similar (intervention impact) aims. I am certain there are other comparable IEA model applications by other researchers and organizations. Nor should one forget the example set by the IEA research tradition as it has influenced OECD-INES (Indicators of National Education Systems), out of which grew the PISA cycles of studies, which today are very keenly followed, and whose publicity worldwide is no doubt greater than that of IEA. Although this may considerably expand the scope of discussion, any such 'extra-curricular' experiences would do justice to the views about IEA contributions to educational evaluation research.

## References

- Bloom, B. S. (Ed.). (1956). *Taxonomy of educational objectives: The classification of educational goals. Handbook I. Cognitive domain.* New York: McKay.
- Bloom, B. S. (1972). Innocence in Education. *School Review*, 80 (3), 1-20.
- Bloom, B. S. (1980). *The State of research on selected alterable variables in education.* Chicago: University of Chicago. Department of Education. MESA Seminar.
- Bloom, B. S., Hastings, J.T. & Madaus, G.F. (1971). *Handbook on Summative and Formative Evaluation of Student Learning.* New York: McGraw-Hill Book Co.

- Bryk A., & Hermanson K. (1994). Observations on the structure, interpretation and use of the education indicator system. In OECD/CERI, *Making education count. Developing and using international indicators*, (pp. 37-53). Paris: Author.
- Coombs, P. H. (1968). *The World education crisis: A Systems analysis*. New York: Oxford University Press.
- Husén, T. (1983). The International Context of Educational Research. *Oxford Review of Education*, 9 (1), 21-29.
- Husén, T., & Kogan M. (Eds.). (1984). *Educational Research and Policy. How Do They Relate?* Oxford: Pergamon Press.
- Husén T., & Postlethwaite T.N. (1994). *The international encyclopedia of education*. Oxford: Pergamon Press.
- Karvonen J. (1974). School democracy and social attitudes of students and teachers in Finnish schools. *Comparative Education Review Special Issue*, 18 (2), 207-216.
- Keeves, J. P. (1972). *Educational environment and student achievement: A multivariate study of the contributions of the home, the school and the peer group to change in mathematics at school*, (Acta Universitatis Stockholmiensis).
- Keeves, J. P. (1995). *The World of school learning. Selected key findings from 35 years of IEA research*. The Hague: The International Association for the Evaluation of Educational Achievement.
- Kuusinen, J. (1967). *Kansainvälinen matematiikan koulusaavutustutkimus*. (International Study of Mathematics, in Finnish). University of Jyväskylä. Publications of the Institute for Educational Research.
- Leimu, K. (1992). Interests and modes in research utilization: The Finnish IEA experience. *Prospects*, XXII, 84, 425-433.
- Leimu, K. (2001a). The Way to a Strategic View on Evaluation. In K. Leimu, P. Linnakylä & J. Välijärvi (Eds.), *Merging national and international interests in educational system evaluation*, (pp. 7-14). Jyväskylä: University of Jyväskylä. Institute for Educational Research.

- Leimu, K. (2001b). Conclusions, Challenges and Visions. In K. Leimu, P. Linnakylä & J. Välijärvi (Eds.), *Merging national and international interests in educational system evaluation*, (pp. 123-153). Jyväskylä: University of Jyväskylä. Institute for Educational Research.
- Leimu K. (2004a). Comparing results from different countries, subjects and grade levels. In J. H. Moskowitz & M. Stephens (Eds.), *Comparing Learning Outcomes. International assessments and education policy*, (pp. 145-177). New York: RoutledgeFalmer.
- Leimu, K. (Ed.). (2004b). *Kansainväliset IEA-tutkimukset Suomi-kuvaava luomassa. Jyväskylän yliopisto: Koulutuksen tutkimuslaitos.* (In Finnish) (International IEA Studies shaping a picture of Finland. University of Jyväskylä. Institute for Educational Research). (425 pages).
- Leimu, K., Linnakylä, P., & Välijärvi J. (Eds.). (2001). *Merging national and international interests in educational system evaluation*. Proceedings of the Conference held at the University of Jyväskylä, Finland on March 19th and 20th, 1998. Jyväskylä: University of Jyväskylä. Institute for Educational Research.
- Moskowitz, J. H. & Stephens M. (Eds.). (2004). *Comparing Learning Outcomes. International assessments and education policy*. New York: RoutledgeFalmer.
- Noonan R. D. (1976). *School Resources, Social Class, and Student Achievement. International Association for the Evaluation of Educational Achievement (IEA) Monograph Studies (5)*. Stockholm: Almqvist & Wiksell International.
- Noonan, R., & Wold, H. (1983). *Evaluating School Systems Using Partial Least Squares*. Evaluation in Education: An International Review Series, 7 (3).
- Owen, E., Stephens, M., Moskowitz, J. H., & Gil, G. (2004). Toward education improvement. The future of international assessment. In J. H. Moskowitz & M. Stephens (Eds.), *Comparing Learning Outcomes. International assessments and education policy*. New York: Routledge Falmer.

- Peaker G. F. (1975). *An Empirical Study of Education in Twenty-One Countries: A Technical Report. International Studies in Evaluation VIII*. Stockholm: Almqvist & Wiksell International.
- Plomp, T. (1992). Conceptualizing a comparative education research framework. *Prospects*, XXII, (3), 1992 (83) Paris: Unesco.
- Postlethwaite T. N., & Husén, T. (1994). *International Encyclopedia of Education*. 2nd Edition. Oxford: Pergamon Press.
- Purves, A., C. & Levine, D. U. (Eds.). (1975). *Educational Policy and International Assessment. Implications of the IEA surveys of achievement*. Berkeley, CA.: McCutchan.
- Stufflebeam, D. L (1976). *Evaluating the Context, Input, Process and Product of Education*. Paper presented at The International Congress of Physical Education, University of Jyväskylä, Finland, 1976.
- Stufflebeam, D. L. (2004). The 21st-Century CIPP Model: Origins, Development, and Use. In M.C. Alkin (Ed.), *Evaluation Roots. Tracing Theorist's Views and Influences*. London: Sage Publications Ltd.
- Super, D. E. (Ed.). (1967). *Towards a Cross-national Model of Educational Achievement in a National Economy*. New York: Teachers College, Columbia University.
- Takala, M. (1963). *Kansainvälinen koulusaavutustutkimus*. [International Study of School Achievement. In Finnish] University of Jyväskylä. Publications of the Institute for Educational Research, 4.
- Weiss, C. H. (1972). *Evaluation research: Methods of assessing program effectiveness*. Englewood Cliffs, NJ: Prentice Hall.
- Weiss, C. H. (1979). The Many Meanings of Research Utilization. *Public Administration Review*, 39, 426-431.
- Weiss, C.H. (1980a). *Social science research and decision-making*. New York: Columbia University Press.
- Weiss, C. H. (1980b). Knowledge Creep and Decision Accretion. *Knowledge*, 1 (3), 381-404.

Participants in the East-West Center seminar, Honolulu, January 1981



**Sitting from the left:**

Mr Shigeru Shimada, Mr. Kimo Leimu, Mr. Russell Cramer, Dr. Torsten Husen,  
Dr. Amarjit Singh., Dr. Harry Passow, Mr. Babu Rrasad.

**Standing from the left:**

Dr. John Brownell, Ms. Seok Hoon Seng, Dr. Zoltan Bathory, Dr. Alan Purves,  
Ms. Soledad Mina, Dr. T.N. Postlethwaite, Dr. Chancha Suvannathat, Dr. Norman Dinges,  
Dr. Hazel Kraemer, Dr. John P. Keeves, Dr. Kathleen Wilson, Prof. Gilbert G. De Landsheere,  
Ms. Christina Rodriguez, Dr. Richard Chadwick, Ms. Anne Hines, Dr. P.B. Welbeck,  
Dr. Eric Casino, Mr. Roy Rhillips.

## CHAPTER 27

# Experiencing Innovation and Capacity Building in IEA Research, 1963-2008

*John Schwille*

In 1973 Harvard University hosted an international conference on the second generation of the IEA's cross-national studies of student achievement (Purves & Levine, 1975). At one of the sessions, Ben Bloom as a founder of IEA told a group of participants (including me) that the conference was a testimonial to being able to do such a large-scale, complex research project without creating another self-perpetuating *bureaucracy*. Bloom's remark undeniably reflected the milieu out of which IEA had sprung—a loose and informal network of very prominent, internationally known educational researchers who decided to innovate and collaborate on an unprecedented venture. It also reflected the culture of the University of Chicago where two of the IEA founders—Ben Bloom and C. Arnold Anderson—held positions of influence. In this culture, the most talented researchers were considered better off if they could pursue their research interests in the absence of organizational constraints. In fact, the culture was one in which the term bureaucracy connoted obstacles and was unlikely to refer to the capacity to do academic work, and especially not research. The irony of the remark was that Bloom was speaking in Longfellow Hall at Harvard in the midst of perhaps the world's richest, most successful and respected self-perpetuating institution, then in its 337th year dealing with the challenges of higher education. IEA in contrast had been put together in somewhat ad hoc fashion to meet the demands of particular studies. But time and institutions change. What would Bloom think of the IEA Data Processing and Research Center in 2010, an organization with over one hundred fifty specialized and technical employees organized to meet standards and maintain quality in IEA studies?



Over the many years I have spent on IEA studies, I have learned much about the creativity it takes to start unprecedented large-scale international research and the capability it takes to complete, sustain and build on these efforts. In principle, innovation and capacity building should be complementary and often they are, but there is also a tension. Innovation calls for capabilities that may be very difficult or even impossible to put together. In contrast, very strong preexisting organizational capability on a large scale can also bring too much emphasis on strictly following established procedures with limited refinements and frequently involves resistance or even hostility to more radical innovation and creativity.

Personally, I was always attracted to the challenges of innovation in research, seeking something new and different that could gradually be made manageable and brought under control. In spite of the advantages of established ways of dealing with such challenges, it was the less conventional that always most attracted me. Fortunately my work was in some of the IEA studies that departed most from the IEA core paradigm. Nevertheless, I have great respect for what has been achieved in the IEA mainstream of mathematics, science and reading literacy, and for the organizational capability that has been developed to make that possible.

Capacity building in the IEA context, in my view, means building up a research consortium's ability to meet six perennial challenges. Although each of the challenges has been more problematical and critical in certain IEA studies than others, no study has escaped a certain amount of struggle with all these aspects:

Identifying appropriate and manageable domains for research

Assembling leadership and support teams with the expertise and other qualities needed for carrying out high quality research

Obtaining the financial resources (principally time and money) needed for research

Building the coalitions necessary for carrying out this research

## Dealing with contextual change

### Institutionalizing the capabilities required to meet all these complex organizational challenges

To be successful, IEA had to achieve a satisfactory resolution of each of these six problems, resolutions that were fragile once they were achieved--hard to maintain and to use in launching subsequent studies. Gradually an IEA template or paradigm with increasingly high standards of quality emerged. The organization started to do repeat studies and to finance itself and this in turn created the resources, institutional memory and capabilities needed for the more innovative and higher risk efforts. The repeat surveys in mathematics-science (TIMSS) and reading literacy (PIRLS) have been the key to building a new IEA, with more resources, more staff, more institutional memory, a bigger repertoire of established techniques and procedures. IEA in 2008 was far from the cottage industry of the 1960s and 1970s. This chapter tells what I have learned over the years about how IEA increased in its capabilities while remaining innovative.

## **From One Subject to Six (1963-74)**

I first learned of IEA in the first doctoral course in comparative education I took from Professor C. Arnold Anderson at the University of Chicago in the fall of 1963. Anderson was a leader in the wave of increased interest in comparative education that marked the 1950s and early 1960s. He was founding director of the University of Chicago Comparative Education Center in 1958. In 1963 he had just finished a term as the fourth president of the Comparative Education Society (now the Comparative and International Education Society or CIES). He was also one of the founders of IEA—one of the researchers who had more or less simultaneously called attention to the need for cross-national studies of educational achievement. Anderson had proposed an IEA type study in his inaugural lecture as the director of the Comparative Education Center in 1958.

In the 1963 course, Anderson called attention to IEA in a statement that I remember as: “The trouble with comparative education is that we don’t have the dependent variables we need for comparative research across countries, but soon we will.” He was referring to the IEA First International Mathematics Study (FIMS) for which data were to be collected in 1964 with results published in 1967 (Husen, 1967). This was the real beginning of IEA, the real beginning of international assessment research as we know it. In this study, IEA demonstrated that it was technically feasible to measure student achievement in mathematics across countries.

I was not directly involved in the first IEA mathematics study. In those years I was totally focused, one could even say obsessed with my own interests, which centered on sociology of education and the study of society and education in France. Still I began to develop an interest in IEA and its leaders, notably Torsten Husen (for more on my memories of Husen and Postlethwaite, see Schwille, 2010).

Arnold Anderson was one of the strongest connections between the University of Chicago Department of Education and Swedish social and educational research. He had been a Fulbright researcher in Sweden in the 1950s and retained a keen interest in many aspects of Swedish policy, society and education and the work of Swedish social scientists like Torsten Husen. He was fascinated in a skeptical way by the social engineering aspects of Sweden. He was especially attracted by the claim that Swedish educational reforms of that period were largely based on empirical educational research. Torsten Husen was an academic leader in this movement, working closely with the Social Democratic government, to bring greater equality and democracy to Swedish education and hence to society as well (Fagerlind, Marklund & Chinapah, 1998; Husen 1983).

Even though I didn’t work on the first mathematics study, I learned enough about the IEA mathematics study in those years for it to have great influence on my thinking. Years later I could see that some of the

criticisms leveled at this study were unfounded. For example, when in the 1980s Iris Rotberg began to publish attacks on IEA studies in various education journals and newspapers, it was clear that some of her assertions were completely wrong. From her citations it appeared that she had never read the primary reports on the IEA studies but had relied instead on secondary and derivative sources. In a rejoinder that colleagues and I published in *Phi Delta Kappan* in 1991, we dealt, for example, as follows with one of Rotberg's misconceptions, namely that the first mathematics study had ignored differences in selectivity of the participating educational systems:

"The differential selectivity or retentivity of secondary school systems, which Rotberg claims was ignored, was in fact a central concern in the early IEA studies... From the beginning, IEA researchers were interested in the statistical problem of how to make appropriate adjustments for the proportion of an age group that remained in school in order to make outcome measures as comparable as possible" (Bradburn, Haertel, Schwille & Torney-Purta, 1991).

### *My Role in the Six-Subject Survey and the First Civic Education Study*

I first started to work on IEA studies in 1972. Since that was toward the end of the Six-Subject Survey, there is much about the early years of that study that I do not know. But when I arrived in Stockholm at the IEA secretariat in 1972, the on-going study was still facing grave problems in dealing with six huge and complex domains at the same time. In moving from just a single subject—mathematics—in its first study of the 1960s to six additional subjects in its second effort of the late 60s and early 70s, IEA overreached. Innovation was the order of the day. The Six-Subject Survey was far more ambitious than the First Mathematics Study had been. It was not a repeat study of mathematics, but instead dealt simultaneously with science, reading comprehension, literature, civic education, French as a foreign language, and English as a foreign language. Twenty-one national centers collected

this data in either 1970-71 or 1971-72 from about 250,000 students, 50,000 teachers and 10,000 schools. From eight to nineteen countries participated in each survey by testing national probability samples of students at one or more of the following levels: 10-year-olds, 14-year olds, and students in the last year of pre-university schooling. The result was 126 separate national samples. The initial analyses were published in 1975 and 1976 in nine different volumes by the Swedish publisher Almqvist & Wiksell, in collaboration with John Wiley & Sons of New York (Caroll, 1975; Comber & Keeves, 1975; Lewis & Massad, 1975; Passow, Noah, Eckstein & Mallea, 1976; Torney, Oppenheim & Farnen, 1975; Peaker, 1975; Purves 1975; Thorndyke, 1975; Walker, 1976). IEA had demonstrated that it could deal with a number of subject matters simultaneously, but it was such a stretch for IEA's fledgling organizational capacity that it has never attempted so many subjects at once again.

The nine volumes did not include the national reports on these studies. In the Six-Subject Survey the national report for the U.S. by Richard Wolf (1977) from Teachers College Press and from Sweden by members of the Swedish team (Hansson, 1975; Husen et al, 1973) were notable early examples of the host of IEA publications that went beyond the international reports.

Back in Chicago in the earlier years of the Six-Subject Survey, I was largely unaware of what was going on in this gigantic study until in 1972 I saw an announcement on a university bulletin board for Spencer Fellowships in Stockholm to analyze IEA data. But when I saw that flyer, I thought this was the chance of a lifetime, and now almost 40 years later, I have not changed my mind. What could be more important than to participate in a research movement to collect data from many countries throughout the world in areas of such significance to future citizens?

When I talked to Anderson to find out more about the fellowships, he suggested that I contact Neville Postlethwaite, the IEA executive direc-

tor, at the upcoming 1972 AERA meetings in Chicago. Screwing up my courage (I was pretty shy in those days), I went to Neville's session and asked him afterwards about applying for the fellowship. He said, "Sure, have a go at it." Not very encouraging, but better than when I asked Judith Torney (the primary leader of the civic education component of the IEA Six-Subject Survey) the same question. She told me I was completely unqualified, having done nothing in political socialization research (and, for that matter, not much in survey research either).

Undeterred, I was able, by asking about the variables in the civic education component, to put together a proposal for secondary analysis of the new civic education data. This proposal must have done the trick, Judith changed her mind. Ironically these inauspicious beginnings evolved into two wonderful relationships—mostly with Judith but also with Neville—that were among the most important of my professional life. I was asked to join the first group of Spencer Fellows in Stockholm as soon as possible. Packing up my unfinished dissertation, I left Chicago for Stockholm in September 1972.

Upon arrival in Stockholm, I met Torsten Husen face to face for the first time. Professor and Director of the Institute of International Education at the University of Stockholm, and Chair of IEA, Torsten was one of the world's most successful and well-known educational researchers, who had an impressive staff of researchers and students to work on the Swedish aspects of the Six-Subject Survey. Torsten himself for the most part did not get involved in the details of the international analyses for the Six-Subject Survey. He left the nitty-gritty of data preparation and analysis to subject-matter authors, Postlethwaite, and the rest of the international team. He irreverently described himself as instead being condemned to carry hat in hand begging for money to fund IEA from all possible sources. Being first among equals in the group of greats who got IEA started, and later duly elected head of the IEA organization, elevated Torsten to iconic status. When the U.S. Academy of Education was founded in 1965, Torsten together

with Jean Piaget and Alexander Luria were the only founding members from outside the U.S. Torsten received an honorary degree from the University of Chicago at about the same time.

However famous he was, when I met him he struck me as warm, approachable and gracious. I found he had all his students, secretary and colleagues call him by his first name – unusual in the U.S. at that time and a custom I always tried to follow in my university life thereafter. After I left Stockholm, I met Torsten infrequently, but always found him in the same warm and gracious mode. Others have told me that Torsten had other less cordial sides, but I always remember him with admiration and pleasure.

Up until the start of the Six-Subject Survey, IEA had little formal framework or organization within which to build capacity. Hardly any organizational capacity was carried over from the first mathematics study to the Six-Subject Survey. The key personnel were largely different. The capability to build on the first study and manage the greater challenges of six new subjects resided in the experience, knowledge and energy of one person – Neville Postlethwaite, who had coordinated the First Mathematics study and who did the first doctoral dissertation based on IEA data (Postlethwaite, 1967; see also Bos & Lehmann, 1995). Fortunately, Neville had the knowledge, energy and leadership ability to launch and become the lynchpin of the operation.

IEA's rebirth as a formal organization rather than just a network of scholars coincided with the beginning of the Six-Subject Survey. After existing informally under the umbrella of UNESCO, IEA became in 1967 an independent nonprofit, nongovernmental scientific organization, incorporated under Belgian law for the purpose of conducting and promoting educational research on an international scale. In 1969 the IEA secretariat and coordination of the Six-Subject Survey moved from the UNESCO Institute in Hamburg to the University of Stockholm. During the Six-Subject Survey, international data processing was shared between the secretariat and a satellite unit at Teachers

College, Columbia University in New York. By the time I arrived in Stockholm, the two data processing units were blaming each other for various shortcomings of the study. As a result, John Hall, who was the leader of the data processing unit in New York, received very little credit for his herculean and largely successful efforts to get the Six-Subject Survey data ready for analysis.

Within this new organizational framework, capacity had to be improvised for six new subject-matters and a host of methodological problems. Throughout the study, demands threatened to exceed capacity in terms of staffing and needed expertise, data processing capabilities, the state of the art in statistical analysis (e.g absence of hierarchical models) as well as ability to report and disseminate data for secondary analysis.

These challenges were met by finding persons of almost superhuman qualities to lead various parts of the study: Neville Postlethwaite, the executive director; Gilbert Peaker, the statistical mastermind of the data analyses; the research directors for each of the six subject areas; the head computer programmers and data managers John Hall in New York and Mats Carlid in Stockholm. But the main burden for bringing this immense enterprise to a successful conclusion fell on Neville's shoulders. So in 1972 when for reasons not shared with me, he accepted a new position at the International Institute of Educational Planning in Paris, this left IEA with a potential crisis. Although there was a small support staff in Stockholm, there was no one left with a researcher's indepth knowledge of the study, or even knowing exactly how to find one's way through all the documentation that had been assembled and kept in notebooks. To fill the gap, Roy Phillipps, who had been the Six-Subject Survey national research coordinator<sup>1</sup> in New Zealand, became the Executive Director. He was totally devoted to IEA and to the success of the project, but his management skills,

---

1. I use this term national research coordinator throughout this chapter although at times in IEA this position was called national project coordinator instead.



though not insignificant, suffered by comparison to Neville. Moreover he lacked Neville's charisma.

I arrived in Stockholm shortly after Neville left. Carolyn Massad, Leigh Burstein and I found that, as new Spencer fellows, we were put on the spot, and likewise the fellows who came a year or so later. The new executive director urgently needed help to keep the study on track. Some of us Spencer fellows analyzed the data for the greater good of getting the main study finished. Others did not. Leigh Burstein, although he became a lifelong advocate and willing advisor/participant in IEA studies, told me that his advisor at Stanford, Lee Cronbach, had advised him to be careful not to harm his own career prospects by becoming too immersed in the morass of Six-Subject Survey data. Several other Spencer Fellows all followed a similar strategy, while nevertheless benefitting IEA in their own way. Dean Nielsen, for example, got a second Ph.D. from the University of Stockholm with a timely thesis and the first secondary analyses of civic education data (Nielsen, 1977).

Carolyn Massad, the Spencer Fellow on leave from ETS, and I had no such hesitation about joining in on the main and most urgent work. But we soon discovered that filling the hole left by Neville was not readily accomplished. For example, by studying the files, I found that Neville had been carrying on a correspondence of epic proportions with national centers, report authors and others. This was long before the Internet, email or even fax, desktop computers, or word processors. I saw for myself what Ken Ross described years later at Neville's memorial service: "Neville had an enormous appetite for hard work—and he always assumed that everybody around him shared the same feeling. He not only worked hard—he also produced huge volumes of high quality work at enormous speed. The energy that he produced was infectious. When he moved into top gear everybody in the vicinity tended to get swept up in the whirlwind—whether they wanted to or not."

We also worked closely but independently with the Swedish IEA team

responsible as a national center for collecting the Swedish data and preparing national reports. Members of that team became friends as well as colleagues. Ingrid Munck, who was a statistician who studied under Joreskog during the early days of the software package for advanced statistical analysis LISREL, became a lifelong friend and collaborator. Her thesis on applications of LISREL to cross-national survey data received IEA's Bruce Chopin award for best use of IEA data in 1985 (Munck, 1979).

During my first year in Stockholm, I did some work on all of the six subjects as a Senior Research Officer. After a year of this, I was asked to stay on a second year to organize a Six-Subject Survey databank to be used for secondary analysis. Nevertheless, my own special interest was in civic education, a component of the Six-Subject Survey in which about 30,000 students, 5,000 teachers, and 1,300 principals responded to instruments having to do with civic knowledge, attitudes, classroom practices and school characteristics. The participating countries included Finland, Ireland, Israel, Italy, the Netherlands, New Zealand, Sweden, West Germany and the USA. Initially my role was to serve as an intermediary between Judith Torney and the brilliant but independent minded team of programmers who worked up the analyses for the civics volume that was to be published by Torney, Oppenheim and Farnen in 1975. Authorship of this volume was already more or less determined before I arrived, but thanks to Judith I got a chance to author the chapter dealing with the regression analysis of the civics cognitive test outcomes (Schwille, 1975). This turned out to be an opportunity to use a technique developed by Newton and Spurrell and to question the standard IEA analyses masterminded by Gilbert Peaker. Gilbert, in the view of Judith and myself, gave undue importance to home background as compared with the characteristics of schooling in explaining differences in scores on all the tests developed for the Six-Subject Survey. Gilbert's point of view was embedded in the analysis model he applied to most of the data. My chapter was one of the few opportunities for a more neutral approach to the question

of how home background versus school characteristics influence outcomes (for an assessment of this chapter in a review of all nine volumes, see Inkeles, 1977).

The first civic education study definitely stood, along with the literature study, on the innovative side of the Six-Subject Survey. In the 1971 data collection, it was the first IEA study to give noncognitive outcomes of education more emphasis than cognitive outcomes. The instruments included not only a test of civic knowledge, but also measures of support for democratic values (including tolerance and support for women's political rights), support for the national and local government, and participation in political activities. Even on the cognitive side, the study yielded important results. In the chapter I was responsible for, we found that stress on rote learning and on patriotic ritual in the classroom tended to be negatively related to civic knowledge and democratic attitudes, while the opportunity to express opinions freely in class had a positive impact. These analyses not only controlled for SES and type of school, but also by means of the Newton and Spurrell analyses, we could show that rote, ritual and open classroom climate had robust effects regardless of the order in which various factors were entered into the regressions. We had such good results with the classroom climate scale that a version of it was again used in the second civic education study of the 1990s. Persons who think there is nothing to be learned about classroom processes from large scale questionnaire surveys should take note.

### *Data Analysis and Data Banking for Six-Subject Survey*

The study that had started with invitation from IEA to social scientists at a conference at Lake Mohonk, New York, in 1967 (Super, 1967) to generate a large pool of potential explanatory variables needed to be put through a statistical template if it were to be kept from falling apart from its own overreach and complexity. Designing this template was the work of Gilbert Peaker, a grand and assertive old man of British statistics, officially retired, who lived in the shadows of

Wordsworth's trees in the north of England and had learned to type only because he needed to send legible epistles to Stockholm to tell the IEA team what to do. This correspondence is filled, not only with the minutia of statistical matrices and multivariate equations, but also with examples and analogies drawn from Romantic poetry, Shakespearean drama and the grand episodes of British history. In spite of the sense of possibility generated by all these references, Gilbert's great contribution took the form of designing one standardized procedure and model so that virtually all independent variables in the Six-Subject Survey could be tested for possible impact on outcomes. This data analysis started with what Gilbert called "scrubbing" which meant partialling out SES and type of school from all the potential explanatory variables. Then the surviving variables were entered into a block regression model which reflected Gilbert's view of how schooling worked and with which, as mentioned earlier, Judith and I did not entirely agree. Our reservations notwithstanding, these procedures did provide a way to bring this gigantic analysis to closure in order to publish the six data analysis volumes more or less on time.

During my first year in Stockholm (1972-73), IEA got a grant from the Ford Foundation to create a data bank for the Six-Subject Survey. This was the first systematic effort by IEA to disseminate its data in a form useful for secondary analysis. My main job the second year was therefore to take charge of the very small team working on this data bank. This job was particularly challenging because the data had not been planned, collected and stored with documentation as a primary consideration. Instead, much of the information needed by secondary researchers was scattered in IEA's Stockholm files and sometimes in places that were difficult for us to find. To prepare for this challenge, Ingrid Munck, Mats Carlid and I set out after the Harvard conference of 1973 to visit major social science data archives in Massachusetts, Michigan and Wisconsin.

We did everything we could to organize and document the data in a way that would be comprehensible to researchers with no previous

knowledge of IEA. We had to reformat the data, choosing to put it in a form that we considered most accessible rather than choosing formats that would be most efficient to use, but more difficult for researchers to manipulate without support from data processing specialists with state of the art expertise.

Fortunately, in planning the data bank, we had visited one of the best data archives for social science research, the Interuniversity Consortium for Political and Social Research, in Ann Arbor, Michigan. So when we needed someone to continue the data bank work after I left Stockholm, we went to them and found Sue Marshall, a very experienced and competent staff member, to come to Stockholm until the data bank was finished (Schwille et al, 1975). In that effort, IEA first demonstrated its ability to document and disseminate a very complex survey data base in a relatively simple form that could be used by secondary analysts.

After returning to the U.S. in fall of 1974, I was gratified when from time to time researchers would tell me that they had been doing secondary analyses of Six-Subject Survey data and had found the data bank relatively easy to use. In one of the most gratifying examples of this feedback, Judit Kadar-Fulop, a member of the Hungarian national center for the Six-Subject Survey, wrote to friends and colleagues about her visit to New York during a three-month trip to the U.S. in 1976: "I had to cope with the task of finding my way in a system completely new to me. It was manageable ... partly because I found the IEA data bank set up there. Thanks to Professor Richard Wolf's generous help and encouragement, I was able to run some additional analyses of the Hungarian IEA data with the double benefit of learning the SPSS system and of testing some hypotheses about the differential effect of school variables in different cultural settings" (letter dated Dec 16, 1976).

## **IEA Reinvents Itself (from Second Mathematics to Third Mathematics-Science and Second Civics, 1975-2001)**

### *The Second Mathematics Study: Disunited but Still Innovative*

After publication of the nine volumes of Six-Subject Survey and dissemination of the data bank to archives, IEA was in a perilous state. Money had been relatively plentiful in the 1960s when the U.S. federal government first started funding educational research and development on a large scale; prominent researchers such as Torsten Husen, Ben Bloom and Robert Thorndyke could get funding for the Six-Subject Survey without too much difficulty. But this was certainly not the case in 1975 when there was widespread skepticism about the need for such studies. Less costly ways of doing international assessment research were being considered. Thus began the most difficult period of IEA history. Even two of the most obvious studies to be repeated – mathematics and science – had a hard time. There was no funding to maintain the sort of centralized secretariat in Stockholm that had been one of the keys to the success of the Six-Subject Survey. These financial difficulties lasted through much of the 1980s. In response, IEA had to rely still more heavily on the in-kind support of its most dedicated researchers and member institutions.

In this context, IEA was becoming a more decentralized organization in which individual member institutions would take the lead as international coordinating centers for specific studies without any headquarters staff servicing all studies. In speaking to the Berlin IEA General Assembly (GA) in 1975, Torsten Husen had advocated giving countries much more autonomy than they had had before. This approach had its advantages, but, in my view, it was also risky in threatening to dilute quality control at the international level. However, this decentralized strategy kept the organization alive until interest in international assessment revived and funding became more available.

At that time the best possibility for keeping IEA alive was the Second International Mathematics Study (SIMS) since countries had already expressed strong interest in it. In spite of this interest, however, SIMS faced critical threats to its survival throughout much of its existence. Once again there was little carryover from the Six-Subject Survey to the second mathematics study in terms of expertise and experience with IEA. Shortage of money was chronic.

Within the U.S. government, the National Institute of Education (NIE) with prime responsibility for educational research was one of the logical places for IEA to seek funding for the second mathematics study. Having finished my dissertation in Stockholm, I took a job at the NIE when I came back from Sweden in 1974. When asked by Torsten Husen and others, NIE considered becoming the national research center for IEA and conducting the proposed mathematics study at least in part in house. I was designated to represent NIE at the IEA General Assembly meeting at the Max Planck Institute in Berlin in 1975 where all the difficulties of launching this new study were discussed. My trip report noted that even though a sufficient number of countries were interested, planning had come to a standstill due to difficulties in getting funds.

NIE never did become the national center for the second mathematics study although it later funded the study to some extent. Instead the University of Illinois became the U.S. national center for that study. Ken Travers, well known as a mathematics educator in the U.S., was chair of the International Steering Committee (ISC). Then the University of Illinois team helped raise much of the funding for the international costs of the study. This gave Americans a lot of influence--too much according to many colleagues from other countries who were not enthusiastic about what was proposed by the Illinois team.

The international coordinating center for the second mathematics study was established at the Department of Education in Wellington, New Zealand with Roy Phillipps as international coordinator. This solution was appealing because the department was willing to pro-

vide substantial resources in kind, especially for international staffing. But Phillipps was stuck in the middle between the Americans represented by Travers and the steering committee and critics represented by Postlethwaite and some of the NRCs. Neville Postlethwaite succeeded Torsten as chair of IEA in 1978, but the chair at that time had very few resources at his disposal. Instead, Neville made good use of his charisma and reputation among the IEA international members to promote his views. The IEA secretariat in Stockholm survived but only in skeletal form until 1987. The IEA national centers were not yet ready to strengthen the organization by paying annual dues and participation fees. In the mid 1970s I had suggested to Roy Phillipps that the organization needed this sort of self-financing and was told it was out of the question.

Of necessity then, capacity had to be rebuilt within the more decentralized organization. For example, IEA studies from the beginning had to gain support from very different constituencies and stakeholders, and this was strikingly evident in the second mathematics study in which it was especially difficult to reconcile the points of view of those who were new to IEA and those who were not. In fact, building a coalition of stakeholders for second mathematics who were united on how to design and execute this study ultimately proved impossible, resulting in a split between countries committed to a true longitudinal study and those wanting to stick with a cross-sectional version.

The need to build broad coalitions drawing together diverse interests which often did not blend well had always seemed inevitable to me in an enterprise like IEA, but I found others were not of the same mind. In 1983 Gary Theisen and colleagues published an article in the *Comparative Education Review* (CER) which became the journal's article of the year. It criticized IEA researchers for lacking a unified conceptual framework or model to guide their analyses. In response, Leigh Burstein and I wrote a response for CER explaining that IEA required coalition building among groups who saw the world of schooling in very different ways (e.g., measurement types, subject matter types,



sociologists etc.) and that they could not be expected to agree on any one framework or analytical model which might reduce the importance of their own stake in any such study (Schwille & Burstein, 1986). Even the domain of mathematics was contested. When leaders of the U.S. mathematics education community were approached in the mid 70s about a possible second IEA mathematics study, they reacted with skepticism and even some hostility. If there was to be a second study, they said, it had to be different from the first study. They complained that even though they had not been involved in the first study, they were put on the spot when the results were headlined in leading newspapers after release of the first results in 1967—news that was interpreted as bad for the U.S. For the first time there was evidence that Japanese young people scored much better than other participating countries—especially the U.S. which ranked 11th out of 12 industrialized countries at the lower secondary level (Husen, 1967). IEA's response was that the first study was not about mathematics per se but rather the use of mathematics achievement as a proxy for a system's ability to produce learning outcomes. This interpretation was not at all satisfactory and even hard to believe among the mathematics educators since the findings of the first study were interpreted in many quarters as a call for changes in the curriculum and teaching of mathematics in the U.S. and other countries. Therefore, in the view of these mathematics educators, if there was to be a second study, the tests had to be designed with strong participation by both mathematics educators and mathematicians to ensure that the tests and results adequately and accurately represented the domains. They wanted a study much more closely linked to issues important in mathematics education and to questions of what counts as effective teaching of mathematics, and how this teaching is linked to learning outcomes.

One of their early moves was a meeting at the University of Illinois in May 1976 to build the coalition needed to get the financial support of the U.S. government. Thirty-five individuals met for five days. In the

end, the tilt toward more emphasis on mathematics as subject matter was accepted and, in the U.S. at least, the mathematics educators largely had their way.

Meanwhile, I left NIE in 1977 to take a job at Michigan State University (MSU), where I continued to collaborate closely with the second math teams. By October 1986 when I attended my last SIMS meeting in Los Angeles, my records indicated that I had participated in more than thirty working meetings on the second mathematics study and the classy Georgian-style student union at the University of Illinois had become one of my favorite places to stay.

In the summer of 1978 I agreed to go to New Zealand for two months to work on planning for the study. To save money, I stayed with Roy Phillipps at his house and went to work at one of the most historic wooden structures of Wellington, which housed the Department of Education. During that time I wrote about the conceptual framework for the study. Some of the positions I had to deal with would sound strange today. One is the assertion that the term “quality” was inappropriate to use as a central concept in an empirical study like an IEA survey. Another, which reflected prevailing views in parts of the educational research establishment, was that curriculum variables should not be prominently featured in the study because they were not policy-relevant.

After returning to East Lansing, Michigan, I went on to organize an additional international planning meeting at MSU in January 1979. Participants included, among others: Ken Travers, Richard (Dick) M. Wolf, Dick Noonan, Roy Phillipps, Bob Garden, William (Bill) Schmidt, Richard Wolfe, Leigh Burstein, Curtis McKnight, Judith Torney-Purta, Carolyn Massad, Andy Porter, Bob Linn, and Dave Berliner. This meeting exemplified one way to cope with IEA’s chronic shortage of funds in those days, namely, getting universities and other agencies to cover the cost of meetings (see Robitaille & Garden, 1989, for list of all the institutions which sponsored meetings, p.viii).

The MSU meeting took place in spite of severe winter conditions. To deal with the weather, the University of Illinois was forced to send out one of its private airplanes to pluck Roy Phillipps and mathematician Izzie Weinsweig out of a Chicago airport. One of my favorite IEA memories is of the two of them, plus the Illinois team, disembarking from the university plane on a very snowy and frigid tarmac in Lansing, Michigan. They probably thought they had landed at the North Pole.

In early 1979, I made one of my most important contributions to IEA (although I certainly did not know it at the time) by recruiting William (Bill) Schmidt, already a full professor at MSU, to go to New Zealand on a trip similar to mine a year earlier. Bill's talents as a sophisticated statistician and very productive researcher were already well known. So not surprisingly, he soon became almost indispensable. After the Second International Mathematics Study was finished, Bill was on leave from 1986 to 1988 as head of the Office of Policy Studies and Program Assessment in the NSF Education Division and soon became a leader in the planning and conduct of TIMSS 1995.

### *The Split between Cross-Sectional and Longitudinal Versions of the Second International Mathematics Study*

A perennial problem in the design of IEA studies is how to handle the fact that any measure of student achievement represents cumulative learning throughout the student's life and throughout his or her school experience. This is not a problem as long as the student achievement scores are reported in isolation from any explanatory variables, but if explanation is desired to analyze the apparent impact of teacher variables, classroom characteristics and other measures that affect learning, one has to find a way to control for earlier achievement. The second mathematics study therefore proposed to do the first IEA longitudinal study at student level with a pretest and posttest to permit computing measures of growth. In the first plans for the study, it was expected that all countries would take this approach, but this design

was resisted by many NRCs who thought the logistical difficulties of collecting longitudinal data outweighed the benefits. Gradually two separate components emerged within the second mathematics study, one a true longitudinal study (Burstein, 1993) and the other a cross-sectional survey (Robitaille & Garden, 1989).

The longitudinal study was strongly promoted by the Illinois center even though many of the NRCs wanted to stick to a cross-sectional approach. This is illustrated by a memo from Ken to Roy and Neville, April 24, 1979: "I do not want to encourage countries to get on the bandwagon for the reduced study, when the International Steering Committee has, it still believes, a much better study (from the point of view of mathematics educators) waiting in the wings."

Two weeks later, Neville responded to Ken in a 7 May 1979 memo pointing out that others were ready to support the more limited cross-sectional approach even if Ken was not: "John Keeves of the [Australian Council for Educational Research] kindly offered to help in the preparation for the cross-sectional part of the study by convening a group of persons to produce the first drafts of the sampling manual and [other] manuals ...."

This split reflected growing tension between the Illinois center and the IEA chair. The original and longitudinal design for the study was seen by some as too much the work of Americans who were overrepresented on the Steering Committee. One of the persons suspected of having too much to say was Leigh Burstein, one of the most active Americans in U.S. IEA circles in those years. Leigh had been part of the first cohort of Spencer Fellows in Stockholm in 1972 where we became fast friends. Later when I took my job at NIE and Leigh started his academic career in the U.S. he was still a presence in my life. We interacted frequently about IEA business and he would call me regularly at my Washington office.

When the second mathematics study started, Leigh became one of the most outspoken members of the Ken Travers kitchen cabinet of North

American researchers who advised him on methodological issues, a group that came to include Richard Wolfe from OISE, Bill Schmidt from MSU, Skip Kifer, another former Stockholm Spencer Fellow from the University of Kentucky and myself. Although there were changes in membership over time, such a group continued to work formally or informally with Bill Schmidt throughout TIMSS and on into the 21st century and IEA's first teacher education study. Richard Houang, Lee Cogan and David Wiley were key additions to the group who are among the most important today. The names of many members of this group can now be seen on numerous publications based on IEA data and co-authored with Bill Schmidt.

Substantively, Leigh's specialty was the development and use of methods for analyzing hierarchical data and he brought this interest to Stockholm already at the end of the Six-Subject Survey, when the lack of such methods was already known to bias the results of regression analyses. Leigh also was the analyst and author who later demonstrated on data from the second mathematics study that Japanese students not only did better on mathematics test scores when the scores were recalculated to reflect the particular curriculum of Japan, they also did better than any of the other countries when Japanese scores were recomputed to match the curriculum of each of the other countries analyzed (Burstein et al., 1993). This analysis forcefully refuted the belief that the Japanese advantage derived from IEA tests that favored their curriculum over the curricula of other countries.

Not all the world was enchanted by Leigh, however, and his relationships with IEA were stormy almost from day one in Stockholm. The problem was that Leigh was outspoken in criticism as in other matters, and these criticisms were often not appreciated. In turn, being passionate and eager to contribute, he resented that he was often ignored and not asked for his advice or invited to participate as much as others. My view is that there were cultural differences involved in all this, and that once one got by Leigh's rough edges, he was a wonderful colleague and friend. This view was widely shared not only among his

colleagues at UCLA but also among the leaders of AERA. Leigh's active involvement in IEA continued until his unexpected and premature death at age 46, when he collapsed while jogging on a morning in July 1994 at a TIMSS meeting in Annapolis, Maryland. His death was a huge shock.

In my view, the Second International Mathematics Study was ultimately successful in spite of being justifiably criticized for taking so long, for sampling problems and for focusing on the specific aspects of mathematics that were of little concern to educational researchers more generally. Data were collected from about 3,900 schools, 6,200 teachers and 124,000 students in more than 20 countries (Robitaille & Garden, 1989; Westbury & Travers, 1990). There was one population of 13-year olds and another of students in the final year of secondary education. Twenty countries (counting parts of Belgium and Canada as countries) were represented in the cross-sectional report (Robitaille & Gardner, 1989). Eight of these countries also did the longitudinal version of the study (Burstein et al., 1993).

The U.S. report on SIMS *The Underachieving Curriculum* (McKnight, 1987) was widely disseminated and discussed—a great success. By April 14, 1987 the report was in its fourth printing with a total of 12,000 copies produced. In a 1987 speech in New York City the U.S. Assistant Secretary of Education for Educational Research and Innovation (OERI), Chester Finn, told members of the IEA General Assembly: “Another example of an international study that tended to ‘remove our blinders’ is the IEA second mathematics study. . . . [It] showed what many had not thought possible—that students can be taught complex mathematics at a relatively early age.”

Even though the second mathematics study was discounted because the growth scores turned out, with some significant exceptions, to be small compared to the margin of error, this study was still able to demonstrate that the difficulties of doing true longitudinal studies cross-nationally could be overcome. It had proved possible to keep

track of and collect data from students and teachers throughout the year in multiple countries (Burstein, 1993). The resulting longitudinal data base remains unique. Years later it was still being used to demonstrate the value of advanced quantitative methods (Muthen, 1996; Wang, 2010).

### *Initial Efforts during the 1980s and Early 90s to Reinvigorate and Change IEA*

Funding prospects for IEA improved dramatically after publication in 1983 of the U.S. report *Nation at Risk* (USDE, 1983). The report was produced by a national Commission on Excellence in Education, which began its work by commissioning papers of interest to the commission. Judith and I wrote one of these papers on the importance of values in U.S. education. Although we knew we were in the minority with respect to the importance of civic education as opposed to science, mathematics etc, we still wanted to make the case for attention to the non-cognitive goals and outcomes of schooling. We had no effect on *Nation at Risk*, but the paper was eventually published in *Comparative Education Review* (Torney-Purta & Schwille, 1986) and then later republished in a volume edited by Epstein and McGinn (2000).

To release the final *Nation at Risk* report, U.S. Secretary of Education Terrell Bell came to MSU where at the same time he gave an award to our MSU Institute for Research on Teaching. Obviously, this was a big event for our college. But of more lasting importance was the fact that the report drew so much of its evidence for problems in American education from the earlier IEA studies. From then on, it was clear that international assessment was back with greatly increased political support and pressure for the continuation of international assessment studies in Washington as well as in other nations.

Shortly thereafter in 1984, I reluctantly reduced my involvement in IEA when I accepted the job of leadership of international research, teaching and outreach in our college as Assistant Dean for International Studies. I was forced to give up the editorship of what

for me was the most interesting volume in the second mathematics study, the one reporting on the longitudinal study. I got Leigh Burstein to take over that volume (Burstein et al., 1993). But I continued my keen interest in IEA and managed to stay involved in smaller ways. For example, in 1985 I accepted Neville's invitation to become a consultant to the IEA technical committee which over time evolved from a rather minor advisory role and gradually became the powerful Technical Executive Group that reviews and passes judgment on the adequacy of methodology in all IEA studies today. I also served on Dick Wolf's U.S. IEA advisory group for some time (for more background on IEA history during the 1980s, see the project reports in the various editions of *IEA Guidebook: Activities, Institutions, and People*).

In the late 1980s a big breakthrough for IEA came in the form of the Reading Literacy Study and the concomitant establishment by Neville Postlethwaite of a data processing unit for this study. I was not at all involved in either of these developments, but mention them here because of their importance for the subsequent history of the studies I did work on. Neville himself stepped in as Reading Literacy coordinator. The data processing unit, unlike earlier units which ended with each study, later evolved into the IEA Data Processing Center (DPC) that is of such importance to IEA today. From the beginning until now, it has been located in Hamburg. Though tiny at first, it streamlined and revolutionized the handling of international assessment data. The following statement from a proposal for the Reading Literacy Study makes clear how modest this beginning of the DPC was.

"The International Coordination and Data Processing Center will be based at IEA Hamburg. Dr. T. Neville Postlethwaite will have the overall responsibility for the international coordination and data processing. Dieter Kotte will act as assistant coordinator and Andreas Schleicher as data manager. Further half time data processors will be hired in the third and fourth year of the project."

This was the first step toward re-establishing centralized coordination



and support services within IEA. IEA had had no executive director from the time Roy Phillipps left Stockholm in the late 1970s until 1990 when, under the leadership of IEA chair Tjeerd Plomp, Bill Loxley was hired as executive director and was instrumental in getting the secretariat re-established in The Hague with funding from Dutch authorities. But it was not until Hans Wagemaker was hired in 1997 that IEA had an executive director who became central to the continuing re-invention of the organization.

From the small beginnings in Hamburg, the secretariat and DPC were gradually built up, and ultimately the DPC became the well-known organization that now employs more than 150 people and is in demand not just for IEA work but for other large scale surveys as well. The key person at the beginning of this development was Andreas Schleicher. People aware of the tensions that have come between IEA and OECD PISA in recent years may not be aware that Schleicher (now head of PISA) was recruited by Neville for international assessment work. He was a brilliant programmer, who designed the software which made international assessment data so much more manageable than in the past and who enabled the 31-nation Reading Literacy Study to be successful with such a small data processing staff (Elley, 1994).

These developments were steps in IEA's efforts to overcome poor quality control and especially problems with sampling in certain countries that had plagued it in its early years. By 1994 a clear agenda for capacity building and improvement had emerged from the IEA Secretariat. This agenda is explained and justified in a paper Schleicher gave at a Stockholm seminar in 1994 titled "Standards for the design and operations in IEA studies" (Schleicher, 1994). In 1993 Andreas had moved from the Reading Literacy Center in Hamburg to the re-established secretariat in the Hague. In his paper Andreas made the following argument for more standardization of IEA's approach in order to have higher standards and more quality control.

“In many past IEA studies, standards have been defined from the perspective of the individual IEA studies, according to their particular needs, and often on an ad-hoc basis. Though, obviously, the adequacy and relevance of standards for the studies are of primary importance, other more general needs are becoming increasingly important. Examples of this are the need to link the results between different studies, the measurement of educational achievement over time, and the growing demand for IEA data by other international organizations.”

“Consequently, there is a need to find agreement on the establishment of a consistent framework of standards at the level of the IEA member countries, in the form of guidelines which can then be implemented by the individual studies according to their needs. These should give countries also clear indications on the requirements and costs of participating in a particular study at its conception, rather than working under a continuously changing framework of ground rules as it is even most recently happening, or even worse, establishing the ground rules post-hoc after the data have already been collected.”

This plan for standards and quality control continued to evolve during the rest of the 1990s. At the 1996 Riga GA meeting Geoff Masters, the GA member from Australia, reported for the Technical Advisory Committee that nine papers were being or had been commissioned on the development of standards in the following areas: (a) test design and construction; (b) questionnaire design and construction; (c) instrument translation, production and administration; (d) population definition and sampling; (e) data quality, entry and cleaning; (f) error identification, containment and reporting; (g) scaling procedures; (h) producing international reports; and (i) making comparisons across IEA studies. Ultimately this work on IEA technical standards was put into a definitive form and published for implementation in a volume authored by Martin, Rust & Adams (1999) (see also Gregory & Martin, 2001).

This work on standards, together with the establishment of the new secretariat in the Hague in 1990, are just two examples of how the

organizational reinvention of IEA was taking off at last in the 1990s under the leadership of IEA Chair Tjeerd Plomp, who had succeeded the late Alan Purves in this role. Purves had lobbied for reforms of IEA in the 1980s (e.g. at the New York General Assembly in 1987), but it took another decade and a half before the reforms became a reality for the organization. Plomp in a 1995 memo asserted that "IEA's reputation as a network of good scholars is no longer sufficient." (GA-36/004.1, Aug 13 1995). IEA was beginning to be described as a "contracting organization," and the need to "professionalize" the organization was emphasized, meaning not relying so much on ad hoc arrangements with individual scholars, universities and other research organizations. For example, after the GA in Indonesia in 1994, the DPC was transferred from the University of Hamburg to IEA itself, with DPC employees becoming IEA employees. IEA faced unrelenting pressure to change in these ways from NSF and NCES, its generous funders in the U.S. government. This pressure was exerted not only directly by the representatives of these agencies when they attended IEA meetings, but also through Gordon Ambach, the IEA General Assembly member from the U.S.

Plomp in a 1995 memo (GA-36/015, Sept 22, 1995) reported that progress that had already been made. A policy of IEA conducting key studies on a regular, predictable timetable had already been adopted even though the details of how often and in what subjects had not yet been agreed upon. Likewise, the obligation of national centers to pay fees was adopted in principle, without yet being completely worked out and implemented. Discussions even got under way with OECD to see if OECD requirements for cross-national educational achievement indicators could be met without creating a parallel structure partly duplicating and competing with IEA. Needless to say today, when IEA and PISA are both in the field doing international assessments of educational achievement, such a congenial arrangement did not prove possible (GA-36/015, 22 Sept 1995; GA-36/004.1, 13 Aug 1995).

### *The Beginnings of the Third International Mathematics and Science Study (TIMSS 1995)*

Once Reading Literacy was completed, these trends led ultimately to the landmark first TIMSS of 1995. I was not directly involved in TIMSS, but MSU played an important role under the leadership of Bill Schmidt. I followed development of the study with interest and also the political context in the U.S. that made it possible. The initial interest in doing mathematics and science together and doing data collection at two points in time came from the U.S. The origin can be traced to the Education Summit held at Charlottesville, Virginia, for the U.S. President and 50 state governors in September 1989. At the end of that meeting, Governor Branstad announced an important agreement on national goals for education:

“In the last two days here, we have made significant progress towards building that national consensus with the leadership of the President and the Governors. In the area of setting national education goals, we unanimously agree that there is a need for the first time in this nation’s history to have specific results-oriented goals. And we’re talking about [goals] . . . [e.g.] in the area of performance of students in international achievement tests in the areas of mathematics and science; . . .” [ from on-line documentation at the George H.W. Bush Presidential Museum].

Then Governor Bill Clinton of Arkansas also played an important role in reaching this agreement. When the exact wording of the goals was agreed upon in the following year, one goal was that the U.S. would become first in mathematics and science by the year 2000. However preposterous it was, that goal was taken seriously enough to warrant the attempt to measure its attainment by an international assessment of mathematics and science at two points in time before the year 2000. The General Assembly in Beijing in 1990 agreed with proposals of the Standing Committee that had been requested by the U.S. for (1) a joint study of mathematics and science and (2) a repeat study later that decade.

With such high level support and visibility, Bill Schmidt's team was able to get a grant from NSF titled Survey of Mathematics and Science Opportunity (SMSO) to develop and validate instruments for TIMSS with research on classrooms in six countries. Schmidt was subsequently designated the U.S. National Research Coordinator for TIMSS, and MSU became the center for an unprecedented in-depth curriculum analysis for the TIMSS countries, based on content analyses of the main mathematics and science textbooks, as well as national syllabi used in TIMSS participating countries (for the publications produced here at MSU, see e.g., Schmidt et al., 1996; Schmidt, McKnight & Raizen, 1997; Schmidt, McKnight, Cogan, Jakwerth & Houang, 1999; Schmidt et al., 2001; Valverde et al., 2002). Funding was never a problem for TIMSS-95, which developed into the largest and perhaps the most complex educational research project ever attempted.

Boston College (BC) took over as the main TIMSS center after Al Beaton, well-known ETS psychometrician, came to BC and was appointed international coordinator in a change of TIMSS leadership. Once Ina Mullis from ETS and Mick Martin (who had been the Irish NRC for TIMSS) joined Beaton full-time at Boston College in 1994, BC became the predominant producer of IEA research. The BC team took primary responsibility for the TIMSS 95 international reports (see e.g. Beaton et al., 1996a,b; Gonzalez & Smith, 1997; Martin & Mullis, 1996; Martin & Kelly, 1996; Martin & Kelly, 1997; Mullis et al., 1997). By the time Beaton retired and the team of Mullis and Martin went on in the late 1990s to become co-directors for TIMSS and PIRLS, BC was launched on an era of productivity unparalleled in IEA history and lasting up until the present time. It was Boston College that was finally able to implement a regular cycle of repeat studies which had been so much talked about since the mid 1980s (see table 1 for the cycle of TIMSS and PIRLS).

The importance of TIMSS in its original and repeat versions would be hard to overestimate. In the spring 2007 issue of *Education Next*, an article on the 13 "most influential education studies of the past

**Table 1.** List of all IEA studies and their role in this chapter, with dates of data collection.

IEA studies that are focus of this chapter	Other IEA studies less discussed or not discussed in this chapter
	Pilot 12-country Study 1959-60
First International Mathematics Study (FIMS) 1964	
Six-Subject Survey (1970-71) , including: First International Science Study Study of Reading Comprehension Study of Literature Education Study of English as a Foreign Language Study of French as a Foreign Language First Study of Civic Education	
Second International Mathematics Study (SIMS) 1980-82	Classroom Environments Study 1981-83
	Second International Science Study (SISS) 1983-84
	Written Composition Study 1985
	Reading Literacy Study 1990-91
	Computers in Education Study 1989-1992
Third International Mathematics and Science Study (TIMSS) 1995	Second Information Technology in Education Study (SITES - Module 1 or M1) 1998-99
	TIMSS-Repeat (R) 1999
	TIMSS Video Study 1995
	Second TIMSS Video Study 1999
Second International Civic Education Study (CIVED) 1999	Pre-Primary Study 1986-1994 (Phase 1); 1989-2003 (Phase 2); 1993-2003 (Phase 3)
	SITES-M2 2000-2001
	Progress in International Reading Literacy Study (PIRLS 2001)

**Table 1.** (continued) List of all IEA studies and their role in this chapter, with dates of data collection.

IEA studies that are focus of this chapter	Other IEA studies less discussed or not discussed in this chapter
	Trends in International Mathematics and Science Study (TIMSS) 2003
	PIRLS 2006
	SITES 2006
First International Teacher Education and Development Study in Mathematics (TEDS-M) – 2007-2008	TIMSS 2007
	TIMSS Advanced 2008
	[Third] International Civic and Citizenship Education Study (ICCS) 2009
	PIRLS 2011
	TIMSS 2011
	International Computer and Information Literacy Study (ICILS) 2013

Source: [www.iea.nl](http://www.iea.nl)

decade” ranked TIMSS second after the U.S. National Assessment of Educational Progress. Already when it was conducted, the first TIMSS of 1995 attracted enormous attention. Perhaps the most unusual and impressive indicator of the policy importance of TIMSS in Washington was that President Clinton himself briefed the press on the TIMSS 1995 fourth grade results in a White House press conference on March 5, 1997. During this whole period, Bill Schmidt not only worked closely with key education policymakers in Washington, but also met directly with many governors and top education officials at the state level. In that era, education had become a very hot issue at both national and state levels. When President Clinton came to Lansing to speak to the Michigan legislature on March 6, 1997, he thanked Schmidt and MSU by name for all this work.

In terms of capacity building, TIMSS 1995 required putting together one of the strongest coalitions in IEA history. A particularly significant

step for maintaining continued widespread interest and support for international assessment in the U.S. was the establishment of the Board on International Comparative Studies of Education (BICSE) at the National Research Council, an arm of the National Academy of Sciences. This step was accompanied by a change in IEA membership for the U.S. Several of us had been unhappy for years about the way IEA was organized in the U.S. with just one university (Teachers College, Columbia) as the national center. One university could not really speak for the U.S. as a whole, and more importantly, could not mobilize the U.S. educational research establishment to take an interest and be involved in IEA studies and secondary analyses as much as possible.

### *The Importance of BICSE to a Changing IEA*

As TIMSS became more imminent, important and visible, the study developed shortcomings that were more and more apparent and in urgent need of fixing. Chester Finn, Assistant Secretary for Educational Research and Innovation (OERI) in the Reagan administration, Dorothy Gilford, a former head of the National Center for Educational Statistics (NCES), and Emerson Elliott, a later head of NCES, took a number of initiatives to solve these problems at national and international levels.

In 1987, instead of returning to the earlier idea that a federal education R&D agency (like NIE in the 1970s) should be the IEA national center. Finn in his speech to the IEA General Assembly in 1987 announced that "the OERI has plans to establish, in a joint effort with the National Science Foundation and the National Academy of Sciences, a consortium of scholars to review IEA (and other international) proposals that seek U.S. participation." Among the responsibilities proposed for this consortium was to designate the U.S. IEA national center and General Assembly member. In my mind, this solution was very much to be preferred over having a government agency as the national center since an IEA center within the government usually does not have all the independence that a research center needs. And no government



agency in the U.S. has the prestige and credibility of the National Academy of Sciences (NAS) in providing research leadership. The NAS zealously guards this credibility by thorough and rigorous review of all its reports and publications as well as all appointments to its boards and committees.

The resulting Board on International Comparative Studies in Education (BICSE) was first convened in 1988. Judith Torney-Purta was one of the founding members. I was invited to join in 1990 and served two terms until 1996. One of BICSE's earliest acts was to designate the Council of Chief State School Officers (CCSO) as the IEA national center and Gordon Ambach, CCSO executive director, as the GA member. CCSO was the U.S.'s version of a council of state ministers of education; it was not in any way a research organization. Nevertheless, this move was a brilliant one. It gave IEA not only a great deal of legitimacy to represent the U.S. since each state's "minister of education" was represented, but it also gave IEA unprecedented access not only to the U.S. Department of Education, but also to the Congress and the various state departments of education, governors, legislatures etc. And Gordon Ambach, having been Commissioner (i.e. minister) of Education in New York, one of the most important states, proved more than up to the task. He took an immediate and long-standing interest in IEA, becoming a member of its Standing Committee, and was responsible for getting many problems with TIMSS ironed out. When he retired, he was selected as one of the few honorary members of the organization.

BICSE also broadened the audience of IEA well beyond the measurement specialists and other selected researchers who previously had been most involved. It always included prominent persons who were vetted for their competence by the NAS National Research Council and who had been not at all active in IEA. Members of the board in those early years included, for example, James Guthrie from Berkeley, Edward Haertel from Stanford, Gaea Leinhardt from the University of Pittsburgh, Francisco Ramirez from Stanford, and Morris Hansen,

Senior Vice-President at Westat. The chairs of BICSE were persons of high status in the U.S. and leaders of educational and social research; they included Norm Bradburn, Director of the National Opinion Research Center (NORC) at the University of Chicago, one of the nation's leading survey firms; Michael Kirst from Stanford and influential co-founder of PACE (Policy Analysis for California Education); Marshall (Mike) Smith, former Dean of Education at Stanford and later to be Under Secretary or Acting Deputy Secretary of Education throughout almost all the Clinton administration; Andy Porter, then director of the Wisconsin Center for Educational Research and later to be Dean of Education at the University of Pennsylvania; and finally Emerson Elliot himself, a long-time government insider and leader in educational statistics and research.

BICSE became an important actor in IEA internationally as well as nationally in those years. Its core mandate was to advise the federal government on whether the US should participate in and fund international studies in education such as those proposed by IEA, and in TIMSS more than ever U.S. funding of international costs was critical since there were no obvious alternatives. BICSE came together three times a year in meetings which were always attended by representatives of NSF and NCES who were paying so much of the international costs of IEA as well as the U.S. national costs. At the very first BICSE meeting in November 1988, the following IEA leaders were scheduled to report: Alan Purves as IEA chair, Neville Postlethwaite for the Second International Science Study, Warwick Elley for Reading Literacy and David Wikert for the Pre-Primary Study. In February 1990, Tjeerd Plomp, the new IEA chair, presented BICSE with a plan for a cycle of repeat studies in key subject matter areas. By May 1991, TIMSS was the biggest topic on the BICSE agenda. Reporting at that meeting were Plomp, David Robitaille as head of the TIMSS international coordinating center; Les McLean from the Ontario Institute for Studies in Education (OISE), chair of the international steering committee for TIMSS 1995; and Bill Schmidt as the U.S. national research

coordinator for TIMSS 1995. Schmidt and Senta Raizen also spoke to the Board about MSU's development project *Survey of Mathematics and Science Opportunity (SMSO)*.

In the early 1990s BICSE even took responsibility for being the contractual pipeline through which federal funding for IEA international costs was transmitted to the IEA secretariat. At the same time BICSE was much involved in the process which eventually shifted the TIMSS International Coordinating Center from the University of British Columbia to Boston College with Al Beaton as the new International Study Director. BICSE's influence was felt outside IEA as well. In 1994 the World Bank and UNESCO joined together to commission BICSE to do an in-depth evaluation of UNESCO's role in international statistics on education. The resulting report was distributed to all delegations at the 20th UNESCO General Conference .

Having a prestigious board like BICSE was also important because in the U.S. this was a period of vociferous criticism of IEA studies as well as support. This criticism included the article by Iris Rotberg referenced above, part of a series which culminated with her attack on international assessment in the prestigious journal *Science* (Dec.1, 1995). The book *Manufactured Crisis* by Dave Berliner and Bruce Biddle (1995) was another prominent example. On the more scholarly side, the thick volume by Robin Alexander (2000) also drew attention to IEA's flaws. And the most judicious and thorough assessment of IEA's short-comings and how they can be dealt with that I know of can be found in the BICSE volume, edited by Andy Porter and Adam Gamoran (2002), on methodological advances in cross-national studies of educational achievement. Nevertheless, thorough assessments of the flaws in IEA studies on a country-by-country and study-by-study basis and how much they are likely to have distorted or undermined the findings are still lacking.

Unfortunately, under the new Bush administration that began in 2001, the U.S. Department of Education decided it no longer needed the

advice of an NAS BICSE and stopped funding the board. The assistant secretary responsible for educational research in the U.S. Department of Education took on the responsibility of appointing the General Assembly member for the U.S., selecting Tom Loveless of the Brookings Institution to act in a more personal capability than had been the case with Ambach and the CSSSO. But by that time BICSE had left its mark not only in its reviews and appointments but also in publishing a number of significant publications relating to international research in education in general and international assessment in particular (e.g. Bradburn & Gilford, 1990; Porter & Gamoran, 2002). It had demonstrated that IEA could be organized nationally in a fashion independent of government and responsive to both the leaders of educational research and government policymakers. I had hoped that this example could be followed in other countries with similar efforts to achieve a better balance between government officials and researchers in IEA at the national level, but this was not to be. IEA continued to move in a different direction with more, not less government influence on its work.

### *Launching the Second Civics Study within a Changing IEA Paradigm*

Although by 1990, IEA had a track record in subjects like mathematics, science and reading as well as plans for improved management and quality control, another of the subject-matters proposed for a second study—civic education—continued to pose challenges in both management and substantive terms. Faced with such management challenges, IEA in the 1990s was increasingly a two-tiered organization—one tier amply funded to improve what had been done earlier and following increasingly strict and controlled versions of the IEA paradigm. This was the case with TIMSS and subsequently PIRLS, the reading literacy study, for which the U.S. government continued to make a generous financial commitment sufficient to do repeat studies in a predictable, high quality and timely way.

The other tier was composed of studies that had little chance of reaching the funding levels enjoyed by TIMSS, PIRLS or OECD PISA (Program for International Student Assessment), but which were thereby freer of organizational constraints than the repeat studies and could therefore strike out in new territory beyond the existing Boston College-IEA paradigm. The second-tier studies would have to continue relying on the productivity, excellence and reputation of prominent individual scholars and/or institutions with reputations sufficient to raise at least minimal funding and attract collaborators and institutions from multiple countries with an interest in participating.

When Judith Torney-Purta proposed a second civic education study in the early nineties, it was soon apparent that this study would have to operate largely in the innovative, but less amply funded tier because there was no funding agency with relevant interests that was likely to provide anything like the resources that were becoming available to TIMSS. For civic education, the key initial funding came from the Pew Foundation in the U.S. (a foundation known for its interest in values) and later from the W.T. Grant Foundation (a foundation with a focus on youth engagement in society). While TIMSS was evolving in the first tier in what was considered a “professionalizing” mode with funding adequate to pay all personnel costs without depending so much on good will and voluntary efforts, civic education was started largely because Judith and those of like mind believed in it. The second civics study under Judith’s leadership became an example of how high quality could still be achieved under the old IEA approach of finding individual researchers with this sort of leadership ability and an intrinsic interest in the subject.

To deal with these challenges and put the survey phase of the second civics study on a more solid organizational footing, IEA set up a competition to select an international coordinating center (ICC). Such a competition was a break with the past, and took place in part because Judith wanted to concentrate on intellectual leadership of a collabora-

tive study and not have to deal with all the other research management tasks (J. Torney-Purta progress report to IEA General Assembly Riga, 1995). Humboldt University in Berlin was the institution selected as ICC and it solicited further funding from the DFG, the German national science foundation. Rainer Lehmann, who had much experience in earlier IEA studies, was chosen as international coordinator. Wolfram Schulz joined the project as associate coordinator a few years later. This German funding and capacity was a godsend; the expertise and skill of Lehmann and Schulz were essential to completion of the study.

Although the initial data processing for civic education was done by the DPC as it had been since the Reading Literacy Study, the DPC was less developed than it is today, and the civics study was more on its own for management than more recent studies have been. For example, at a 1999 meeting of the steering committee for the study, we had to decide for ourselves how to report data for countries that did not meet our sampling requirements whereas subsequently the TEG and sampling team began to control this reporting. Nevertheless, the TEG, then chaired by Ray Adams, was already demanding that sophisticated Item Response Theory (IRT) scaling be used for scoring and reporting all attitudinal scales in the civics study. This had not been done before and resulted in puzzlement among political scientists who wanted to use data from the study. As standards were made more rigorous and explicit, they would become harder to meet for the less well funded studies.

### *Challenges of the First Phase of the Second Civics Study*

The substantive challenges of the second civics study were many. Not only were there skeptics of the earlier civics cognitive test and the reliance on attitude scales in the First Civic Education study, but the landscape in which civic education was situated had changed. These changes in landscape began to take shape in the period of student rebellions in the 1960s which shook the foundations of civic education. The field continued to evolve in response to breakthroughs in attitudes

about gender and politics, substantial changes in the political climate and regimes of participating countries, as well as international hostilities and economic crises, all contributing to greater and greater globalization and culminating in the breakup of the Soviet Union and its client states. The result was an era in which the ideologies of markets and multiparty democracies were at least temporarily dominant across the world.

Apart from these changes, design of the first study also needed reconsideration because the distinctive possibilities of civic education in the first study had been constrained by decisions made to apply to all the six subjects (surveyed together within a two-year span of time in the 1970s). Twenty years later it was clear that the goal in the civics study was even more ambitious than in the studies of mathematics, science etc. It was ultimately to lay a firmer foundation among young people for democracy, including human rights and understanding among nations. One of our most enthusiastic boosters and collaborators, Ingrid Munck, an experienced IEA hand since the Six-Subject Survey, made sure we were keeping democracy front and center as the ultimate goal.

In the end, the second civic education study overcame these challenges successfully with more than two dozen countries participating (11 post-Communist countries, 12 other European countries, 2 Latin American countries, Australia, Hong Kong SAR, and the United States). Approximately 90,000 fourteen-year-old students from 28 countries took the test in 1999 as well as a questionnaire, each instrument about 40 minutes long. Data was also collected from teachers and principals. As a result by the time the second civics study ended, comparative research on civic education had established itself as a major and continuing area of interest to policy-makers, researchers and educators alike, all looking for more precise and valid ways of understanding the conceptions, attitudes, knowledge and actions of young people on whose learning our civic futures will so much depend (Steiner-Khamsi, Torney-Purta & Schwille, 2002; Torney-Purta,

Amadeo & Schwille 2003, 2010; Torney-Purta, Lehmann, Oswald & Schulz, 2001; Torney-Purta, Schwille & Amadeo, 1999). In short, this study demonstrated IEA's ability to deal with many complexities of outcomes and explanatory factors that had not been addressed in studies of mathematics and science.

To explain this success requires a more detailed account of the study's history. The Second International Civic Education Study began to get off the ground when Judith Torney-Purta was asked to prepare a proposal for a second civic education study for the General Assembly meeting of 1993 in Spain. It was during this meeting that Judith had the idea that the study should be conducted in two major phases, one to understand the nature and context of civic education in different countries and the other to design and carry out large-scale surveys of schools, teachers and students with tests of civic education knowledge similar to the other IEA studies. When Judith asked me to join the study shortly thereafter, I gladly became a member of the planning (later steering) committee to work under her leadership.

Judith is an exemplary research leader from whom we all learned a great deal. She truly believes in a collaborative approach. She is efficient in meeting high standards, while maintaining her commitment to the well-being and interests of others. These personal qualities are important in explaining how she was able to make meetings productive in IEA and to manage a very complicated project with very limited resources, collaborating with people from different backgrounds who were not paid extra for this work. For example, she knew when to stand firm, when to compromise, when to start early, when to quit, when to work hard, when to have fun, what to decide now, what to leave until later. These matters of judgment show how in a week or so of meeting with people from over twenty countries, we could reach agreement on highly complex and even contentious matters of study design and instrument development, while keeping in mind the underlying policy and practice issues that the study was meant to address. Not all IEA studies have been as successful in doing this.



Likewise after participants went home from the meeting, Judith found ways of assigning tasks, delegating responsibilities, and keeping track of many different tasks at the same time. As technology developed, she was quick to take advantage of it in enhancing her leadership skills. Within the limits of her funding, she got very good help by recruiting and mentoring highly competent doctoral students, including most notably Jo-Ann Amadeo, who continued to collaborate closely with Judith after receiving her Ph.D. In short, Judith's title, chair of the steering committee, in no way captured the multifaceted nature of her leadership.

Although in the second civics study, I worked on many memos, guidelines and instruments for the study as well as on reporting results, my passion was again to deal with some of the particular challenges that had either not been recognized and/or resisted solution. Two of these were (1) how to deal with competing points of view among educators and constituencies on the very nature of civic education and (2) what to do about the nationally specific aspects of what students were learning that were not taken into account by the way IEA had heretofore developed its tests.

Concerning competing points of view, I was worried that, if the usual IEA paradigm were followed, the intended curriculum analyses would be based on official Ministry of Education policy and curriculum documents, neglecting many other factors that impinge on the civic education curriculum actually offered and experienced in school. In my view, this would have been problematical in two respects: the official position would not be likely to give sufficient attention to all the points of view about civic education which were important in the country and which could potentially influence educators and future governments. And in giving the appearance of consensus to official documents, many of the critical specifics of civic education would be lost in generalities and rhetoric.

In a 1997 AERA paper (Schwille, 1997) to illustrate the differences with

which we would have to deal, I described two imaginary teachers in different countries in order to make clear how different teachers' views and positions on civic education can be. One was a male teacher in his 50s (whom I called Mr. Homeland) who had been teaching history in an Eastern European country for almost 30 years. Much of this time was spent under a Communist regime where the teacher had to be very careful when speaking to students about political matters. Later with the Communists out of power, he was able to put more emphasis on the long historical struggle of his country's predominant ethnic group to achieve autonomy and maintain a distinctive national identity. He therefore favored measures to purify the national language of foreign influences and was opposed to government schools giving too much emphasis to the languages or traditions of relatively small minority groups even though they had lived in the country for generations. He also favored continued teaching of the dominant religion of the country in school, believing that this was a more effective way of teaching civic virtue than a secularized form of civic education. Finally, he considered his students materialistic and overly cynical not only about politics but also about the national traditions he cared so deeply about.

The second teacher was a female, 38 years old, who had also been teaching history, but in a West European country (I called her Ms. Diversity). She opposed many of the policies of her country and was not at all in sympathy with the tendency of citizens, expressed in polls, to be less supportive of the rights of immigrants than they had been earlier. In particular, she thought that many Muslim children in the country had been poorly treated. Instead of teaching any one religion in schools, she favored comparative studies of the world's great religions. She was also a strong advocate of more European unity. She believed that people could identify with their local setting, their nation and Europe at the same time. Like Mr. Homeland, however, she feared the influence of multinational business interests in the movement toward European integration. But whatever she believed, this teacher

was careful not to mention her active involvement in politics in her classroom since any expression of partisan political ideas was strictly forbidden for teachers in her country.

These two vignettes illustrated our intent in the second civics study to take into account multiple and often competing points of view. In the first civics study of the early seventies, this was not possible. That was a period in which the Cold War was still in full swing and in which the ideological differences between countries were pronounced. In particular, there was much open opposition to the principle of multiparty, representative democracy as a desirable goal for countries throughout the world. The very idea of a civics study at that time was controversial and in many countries posed too many difficulties to be carried out. The resulting first civics study was limited to a self-selected, very restricted set of countries who shared many of the same political traditions characteristic of multiparty representative democracies, namely, West Germany, Finland, Ireland, Israel, Italy, Netherlands, New Zealand, Sweden and the United States.

The second civics study was designed to be more responsive to distinctive national traditions and competing points of view. To begin this process, 18 framing questions were formulated that would encourage countries to study and express insofar as possible a full range of views and issues concerning civic education in a particular country, however exceptional or distinctive that country might be. The intent was to avoid the bias that in the first civics study resulted in such a narrow range of countries. The following are a subset of these questions:

If "democracy" is a central concept, what does it mean within the national context and what are young people expected or likely to learn about it by age 14 or 15?

What are young people expected or likely to learn about law and the rule of law, the constitution (written or unwritten), the courts, the national/regional legislature, elections and other institutions of government by 14 or 15?

What are young people of 14 or 15 expected or likely to know and believe about dissent or protest as a way of changing government policy?

To what extent are young people expected or likely to have learned by 14 or 15 that economic principles (such as free market principles vs. state intervention and control over the provision of goods and services) are connected with government or political issues?

What are young people expected or likely to have learned by age 14 or 15 about those belonging to 'minority groups' or other groups which see themselves as disadvantaged or disenfranchised (as defined by ethnicity, race, immigrant status or other characteristics) in relation to the rights and obligations of citizenship?

What are 14 or 15 year olds expected or likely to have learned about relations between their country and other countries?

It took a lot of discussion and debate to avoid questions which were phrased in a manner that might discourage or restrict the free expression of the most significant voices and differences of opinion at the national level.

For the preparation and vetting of responses to the framing questions, each country was asked to form an expert panel representing the various points of view most important within the country. Already by 1997 when I gave the AERA talk it was clear that this ideal had been realized more completely in certain countries than in others. In the U.S. national report, for example, there was no explanation or even mention of the fact that there had been a huge national uproar over proposed national standards for the teaching of history earlier in the decade, an uproar in which issues of civic education were central to the dispute.

In the case of a highly conflicted country where it was impossible to arrange for persons with opposing points of view to work in the same

room on the same panel, alternative arrangements were proposed, such as allowing for written expression of opposing points of view. Although not as representative as we had hoped, the panels played an important role in making sure the case studies would be more representative than the analysis of a single researcher or cohesive team of researchers might be.

In addition to responding to the framing questions, each country was asked to prepare a national case study of civic education following international guidelines from the steering committee. In order to find out how the country planned to conduct its national case study, each country had to submit a proposal in response to the guidelines. As far as I know, the submission and peer review of proposals has never been part of the IEA approach to such case studies, but it seemed necessary in this study in order to make clear that qualitative case studies can and should benefit from careful design as much as quantitative surveys, with special data collection to increase the rigor of the studies and to avoid the “armchair” approach of one presumed expert sitting down and writing a case study without the benefit of any added data collection or even consultation with other experts.

In addition, the responses to the framing questions were used by the international steering committee to formulate three core questions as the basis for further design and analysis. The three core questions focused on (a) learning about democracy, (b) development of a sense of national identity or national loyalty and (c) learning about social diversity and social cohesion. There was a great deal of agreement on the centrality of these issues across countries, but with the expectation of great differences of opinion between and within countries on what to do about them. Countries responded to these three questions with details on official curriculum coverage, pertinent examination questions, typical class activities and assignments, extracurricular and out-of-school activities, and use of the media regarding the core topic in question. The intent was an enlarged curriculum analysis covering not only official curriculum guides and textbooks, but also concomitant

sources of information that might or might not be congruent with what was officially prescribed.

The first meeting of the national research coordinators for the Second Civic Education Study was held in Enschede in the Netherlands on July 17-20, 1995, with about 30 participants (Torney-Purta & Schwille, 2011). Among other things, the national case study proposals that had been submitted were discussed. Training was given in how to organize and conduct focus groups inasmuch as this was a method of data collection recommended for the case studies. It became clear during this meeting that some of the Eastern Europeans felt that the study did not yet represent them and their region adequately. I remember being upbraided by one of the authors of a national case study for presuming that I knew enough about the country in question to review its case study proposal. But gradually under Judith's leadership and with help from two distinguished Polish colleagues--Barbara Malak and Barbara Franzak-Rudnicka--the ambiance at the meeting became more positive.

Working in this fashion, by the end of the study's first phase, reflecting three years of effort, we had a major publication in which each of the national research coordinators had a chapter based on their national case study (Torney-Purta, Schwille & Amadeo, 1999). The book was nicknamed "the brick" because of its length. "The brick" was a success, named an Outstanding Academic Title of 2000 by the journal *Choice*, which is used by bibliographers in research libraries to select books. In 2000, about 600 books received this award from the 6700 titles reviewed and the 25,000 volumes submitted to this journal. To our knowledge, this was the only major book award ever received by an IEA international publication. Later in a cross-national analysis of the same cases published under Gita Steiner-Khamsi's leadership, Jo-Ann Amadeo and I did a chapter on the role of schooling in civic education (Schwille & Amadeo, 2002).

### ***Challenges of the Second Phase of the Second Civics Study***

In the first phase being responsive to national diversity was the

essence of our plan for the case studies, but in phase 2, the more traditional survey phase of the study, it was not at all clear how to take national specificity into account. The IEA paradigm, in general, favored methods for arriving at commonalities and consensus and not differences that were so specific that they were hard to synthesize across countries. But in the case of civic education, limiting the study to a set of standardized instruments focused on the common denominators of civic education did not seem sufficient -- at least to me. For example, the study of history, which is often the largest part of civic education in school, concentrates on specific aspects of national history and not on cross-national generalities. It is not possible to assess what students learn from this sort of civic education without test items that are specific to the country in question. These items in turn would not be appropriate in other countries where students have studied for the most part a completely different historical narrative.

Thus, for the civics study we initially planned to require countries to use a certain number of nationally specific items to include in the cognitive knowledge part of the instruments. But to do so, we would need appropriate ways of reporting on these national items in the international report. We wanted to know if we could use scores based on their national items as part of profiles, not to make cross-national comparisons, but to illustrate the characteristics and qualitative analyses of particular countries. Or alternatively, would it be possible use such items to make cross-national comparisons (e.g. through calibration to the cross-nationally standardized parts of the tests)? In the end, however, resources and time did not permit following through on these ideas.

When the responsibility for writing cognitive items was assigned, there were inadequate funds to pay specialized item writers, so all the members of what was then called the Planning Committee were asked to work on them. Early in the study I had suggested to Judith that, since the cognitive test in the first civic education study had not been of high quality, perhaps a high quality civics test was not possible. Given that the noncognitive outcomes could be considered more important, we

could just dispense with a cognitive test so as to concentrate on the rest. The response was “No way!” A cognitive test was essential and a part of IEA’s standard template or paradigm. And not long after making this suggestion, I found myself along with the rest of the steering committee writing items for the cognitive test. Afterwards I teased Judith by saying that anyone who questioned whether something she asked for could be done would soon find him or herself doing it. Actually, I enjoyed the challenge of trying to come up with items that would not be subject to the criticisms that I and others had raised.

In general my responsibilities for the later phases of the second civics study were much less than in the first phase, since I was heavily involved in African work and also starting to turn my attention to planning an international study of teacher education. Judith, however, to her credit had no intentions of just handing over the two international reports to be published and leaving it at that. With her passion for the understanding and improvement of civic education and especially for its emphasis on democratic values and human rights, she was determined to get the message of the study out to as many audiences as possible, to continue to do secondary analyses so that the full potential of the data could be realized, and to help initiate others, especially young scholars, into research in this field using data from the second civics study (see bibliography of her chapter in this volume).

### **The First Teacher Education Study: IEA Moves into Higher Education**

#### *IEA’s Increased Capabilities Face a New Challenge*

My final IEA experience was to participate in something unprecedented—an IEA study of teacher education. TIMSS was a model that worked for K-12, but what about higher education – how much of the old paradigm would fit? The six organizational challenges mentioned in the introduction were back in force. We started without any tailor-



made model for a cross-national study of teacher education curricula, organization, processes and outcomes. We needed to figure out the domain for the research, what sort of expertise was needed, where we could get funding, and how to build new coalitions--all in a context that differed greatly from that of elementary and secondary education in which IEA had heretofore worked. We had to learn more about how to collect data from national probability samples of institutions, teaching staff and students in various types of institutions, from comprehensive universities to teacher colleges. In short, the IEA Teacher Education and Development Study - Mathematics (TEDS-M) was a first. It was the first in all of higher education, not just in teacher education, to do an international assessment of learning outcomes based on national samples. Needless to say, I was eager to be involved in such a pioneering venture, both because of my prior IEA experience but also because teacher education was one of MSU's primary strengths and an area in which I had been involved since arriving at MSU in 1977 (see, e.g. acknowledgements in Schwille & Dembele, 2007).

By 2002, IEA was a stronger organization than it had ever been before. Under the leadership of Hans Wagemaker, the organization had a coherent strategy for its research which was much more sustainable than in the past (for recent examples of completed studies, see, e.g., Martin, Mullis & Kennedy, 2007; Martin, Mullis & Foy, 2008; Mullis, Martin & Foy, 2008; Mullis, Martin, Kennedy & Foy, 2007). The strength of this strategy was built around the success of the repeat studies in science-mathematics (TIMSS) and reading literacy (PIRLS). These studies embodied a capacity to do projects on time with higher quality standards (as documented in the IEA standards document, Martin, Rust & Adams, 1999). In turn, the participating countries paid participation fees which gave the organization much more viability and even some financial reserves. The Data Processing Center and Secretariat were becoming staffed at a size and level of relevant experience that would have been unimaginable at earlier times. It was from this position of strength that IEA decided to take on the challenge of a teacher education study.

By this time each of the elements of IEA organization were run in a more business-like manner. The governing body continued to be the General Assembly, composed of one representative of each participating institution, typically the director. But this GA was much different than it was in the early days. There was less participation by eminent researchers and more participation by Ministry of Education types who could, when needed, speak for their governments and thereby help assure access and support for IEA studies. This was good for getting support for studies, for developing organizational capacity more generally and for building the necessary coalitions. It was perhaps not so good for the creativity and innovation that research calls for.

The meetings of this new General Assembly were carefully prepared so that the GA could systematically review existing studies and especially consider whether to approve new ones. Preparation was in the hands of the Executive Director and the Secretariat, notably Barbara Malak who had first become known to IEA in the early days of the second civics study as a Polish social psychologist whose exceptional qualifications for civic education were rooted not only in her research credentials but tested in the fire of the Polish Solidarity movement and who, when later hired to work in the Secretariat, became a key person in the running of IEA.

At the General Assembly meetings, the IEA Standing Committee and Technical Executive Group (TEG) played special roles in bringing IEA work under closer and more centralized scrutiny. These committees met with all study directors in the days before GA plenary meetings. The TEG was especially important, no longer as an advisory committee, but as a panel empowered to offer methodological leadership that would be binding on participating studies. In the execution of studies, the IEA Data Processing Center in Hamburg also played an increasingly important role, especially in the preparation of manuals, in the increasingly sophisticated initial processing of data for studies, and in the preparation of databanks for secondary analyses. DPC personnel assigned to studies were by now expected to attend all the important

meetings of each study (e.g. national research coordinator meetings). Since Wagemaker and Malak also attended many such meetings, the contributions of these leaders from the central IEA offices and services were a major advance in bringing institutional memory and capacity to bear on new projects.

By 2002 the repeat studies TIMSS and PIRLS organized by Boston College, the DPC and Secretariat were being referenced by IEA staff as the model from which new studies would be organized. Each year that I attended the GA from 2002 to 2010, Mullis and Martin reported in clear and polished fashion on the vast amount of high quality work accomplished that year on TIMSS and PIRLS. Their powerpoints and explanations were virtuoso displays of research competence and mastery of the management challenges of these gigantic and highly technical studies. All these elements of the current IEA capabilities had been in place by the end of the second civics study, but in a less developed, regulated and standardized form.

Although the intent was for the first IEA Teacher Education Study (TEDS-M) to operate in the business-like manner of TIMSS and PIRLS, circumstances conspired to prevent this from happening consistently. Organizational interests, funding opportunities, and researcher points of view, all under the strain of an unprecedented study, were far too different to allow for the more highly integrated approach of TIMSS and PIRLS at Boston College. Thus, the critical organizational challenge of this study was how far to stick to and how far to depart from the existing Boston College-IEA paradigm. Thinking in terms of intended, implemented and attained curriculum was relevant to higher education, but how to put this framework into practice within a completely new context, that of postsecondary education, was not clear-cut. Within the postsecondary system, there was more autonomy for institutions to establish different programs as well as to cooperate or not cooperate with the study. Within higher education institutions, students (having become adults) had much more autonomy, and instructors had much more autonomy than in elementary and second-

ary school. These challenges led to heated arguments within the TEDS-M team over, e.g., how to deal with consecutive programs in which future teachers learn mathematics in a first university degree and pedagogy in a second, whether it was possible to study teacher education without collecting data on what new teachers were learning on the job, whether it was sufficient to limit outcome data to student teachers' knowledge of subject-matter and pedagogy, how to get a fix on opportunities to learn (including those not documented in syllabi and other curriculum materials), and what curriculum information to collect. An even more basic question which troubled the team almost to the end of the study was to agree on the appropriate unit for analyzing and reporting results while taking into account the great diversity of these teacher education systems.

### *From Pre-TEDS to TEDS-M*

At Michigan State, we had our own strongly held ideas, based more on our own earlier research on teacher education than on IEA research. In fact, we started to work on TEDS on our own before it became an IEA study. The origins of MSU's intent to conduct international research in teacher education go back to 1985 when, under Lynn Paine's leadership, we proposed such a study for our first National Center for Research on Teacher Education. Unfortunately, the U.S. Department of Education was not ready to support international research as part of that center. Keeping this in mind, 15 years later in spring 2000, I convened a group of faculty at MSU (namely Teresa Tatto, Bill Schmidt, Lynn Paine, Chris Wheeler, Bob Floden and Gary Sykes) to discuss what MSU's next priorities should be for international research in education. At the time, our role in TIMSS of the 1990s had almost come to an end. I was afraid we might lose the exceptional capabilities we had built up for doing international research. We discussed this and surprisingly, quickly came to a consensus. It was time to do international research on teacher education—TIMSS and other studies had made clear the importance of well qualified teachers, but there was no cross-

national data to help answer the question of whether teacher education does or does not help teachers become effective. For such research MSU had unique relevant experience, for example, Teresa Tatto's pioneering study of the cost-effectiveness of three different teacher education programs in Sri Lanka—still the best quantitative study of teacher education in a developing country, in my view (Tatto et al, 1993) and Lynn Paine's leadership in cross-national case studies of teacher mentoring and induction (e.g. Britton, Paine, Pimm & Raizen, 2003). This intent to do more international research on teacher education is documented in emails I sent out to these people in the years 2000 and 2001. For example:

“Tomorrow Tuesday, October 17 [2000] we will have another meeting to continue discussions of last spring on what directions the college should take in international research related to quality of teaching and teacher learning. The idea is to consider how to capitalize on our tremendous comparative advantage in this area based on our track record with TIMSS, teacher mentoring-induction research and various related projects in developing countries. Ideas run the gamut from initiating a new large-scale IEA cross-national study of teacher quality and development to smaller more focused studies to follow up on particular issues raised by our previous research.”

At one of our meetings, Teresa and Bill agreed to head up writing a proposal to get NSF support for an international study of teacher education. This was the beginning of the MSU study we first called Pre-TEDS (now called MT21). The proposal, first drafted by Teresa in November 2000, was submitted to NSF as a preliminary proposal in April 2001 and as a full proposal in June of that year. Although submitted independently of IEA, the proposal suggested that the proposed project might subsequently seek to become an IEA project. As is often the case, since the first proposal was reviewed with some favor but not funded, we submitted a second full proposal to NSF in November 2001, one year after Teresa drafted the first version. It took NSF so long to review that second proposal that by summer 2002, we assumed it was dead.

At the same time, unbeknownst to MSU, IEA had been preparing on its own to launch a teacher education study. In spring 2002, the secretariat announced a competition for an institution to take the lead in that study. At MSU, we decided immediately to compete and so apparently did the Australian Council for Educational Research (ACER). Lawrence Ingvarson called Gary Sykes at MSU to see if we were interested in participating under ACER leadership. We said, "No," we would submit on our own. In reviewing the submissions, IEA surprisingly decided that they wanted both MSU and ACER to work together to design and coordinate the study. Initially, there were two main tasks: prepare a preliminary proposal to submit to NSF in September 2002 and a more extensive proposal for discussion by the IEA General Assembly at its Morocco meeting in October of that year.

The ACER-MSU partnership was what in American parlance is called a "shot-gun" marriage, an alliance which is not the free choice of either partner. As in many such marriages, this one was marked from the beginning with disagreements, especially since this marriage of two very different institutions was to be consummated by seven leaders (plus other influential collaborators): three from MSU (Teresa Tatto, Sharon Senk and myself), three from ACER (Lawrence Ingvarson, Ray Peck and Glenn Rowley – replacing Adrien Beavis), and finally Hans Wagemaker (supported by indispensable IEA staff, notably Barbara Malak, Sabine Meinck, Ralph Carstens, Falk Brese as well as Jean Dumais, our extraordinarily competent sampling and survey expert from Statistics Canada). As the project progressed, persons outside the core group of seven and IEA staff became increasingly important as well, e.g. Mark Reckase, a very distinguished MSU psychometrician who had been a vice president of a major testing firm (ACT) and later a vice president of AERA as well; Kiril Bankov, the Bulgarian mathematician who had worked on TIMSS; Michael Rodriguez, a psychometrician at the University of Minnesota who got his Ph.D. at MSU; and Inese Berzina-Pitcher, the extraordinarily competent TEDS consortium coordinator at MSU, originally from Latvia.

In the early stages, there was much tension and confusion about who, among this leadership team, was really in charge. In addition, there was much lack of understanding between MSU and ACER. Ultimately, Hans and the NSF project officer attempted to resolve this problem by making Teresa Tatto the principal decision-maker. But that was not sufficient; although the seven leaders constituted a Joint Management Committee (JMC) which met and dealt with the tensions to some extent, the JMC never became a body able to resolve all disagreements as was initially intended.

In the early days of MSU-ACER collaboration in 2002, we learned to our astonishment that NSF had decided to fund the other, earlier MSU proposal we considered dead. Since at the time there was no external funding for TEDS, this grant was a lifesaver. But it also raised further concerns about relationships among MSU, ACER, IEA, and even within the MSU team. It was the source of strong tensions to follow. To try to resolve these tensions, the JMC met in Melbourne in February 2003 when it was decided that this first MSU grant would become an integral part of TEDS (hence the acronym P-TEDS meaning Pre-TEDS). This agreement would help ensure that the larger study would benefit more fully from the work of P-TEDS. Although this decision only partially reduced the tensions, the pragmatic and urgent need to make and carry out decisions about the design and execution of the study overcame many disagreements, especially since all the leaders were unquestionably committed to bringing the study to a successful conclusion. Teresa Tatto deserves special credit for persevering and working unceasingly and incredibly hard in the face of much criticism and opposition.

The future of the larger TEDS study was still in doubt until we could get a grant large enough to cover much of the international costs of the study. NSF was the only funding agency we knew of that was at all likely to provide that funding, and MSU had therefore to make the case for another NSF grant. Unfortunately, NSF was very hesitant about investing such a large amount in such an unprecedented and large-scale study; they would have to justify this to the U.S. commu-

nities of mathematicians and mathematics educators as well as other key NSF officials and constituencies. Addressing and meeting their conditions took much time: submission of the preliminary proposal in September 2002, the first full proposal in February 2003, and submission of the second full proposal not until almost two years later in December 2004. One example of why it took so much time was the addition of Sharon Senk to the leadership team. NSF made clear that one of the principal investigators (PIs) had to be a mathematician or mathematics educator, and neither Teresa nor I met that stipulation. But getting one of MSU's well-known mathematics educators to make such a commitment of time and reputation to a study outside their primary area of work was not easy. We had to be patient and were ultimately rewarded by persuading Sharon Senk to fill this role. She was well-known and respected for her other work, especially for co-directing with Zalman Usiskan the development of secondary school materials for the University of Chicago School Mathematics Project.

It was during a snowstorm two days before Christmas 2004 that the final proposal went in. Teresa and I trudged through a lot of new fallen snow to the MSU Contracts and Grants Office to make sure the proposal got to NSF before Christmas. But again NSF was in no hurry. After waiting for the final questions from them until July 2005, we did not get the grant until September 2005. One of the final obstacles was NSF's insistence that we have funding for U.S. participation as well as funding for international costs. This demand was very reasonable, we expected it, but unfortunately we had no such funding at that time. Finally, NSF compromised and decided to begin the grant, provided that we agreed to come up with the U.S. national funding within a year. In my view, the only way we could get this funding was to rely on Bill Schmidt's reputation and demonstrated ability to obtain large grants. Bill agreed to try. And by the next year, using what had already been accomplished in P-TEDS (as well as his other accomplishments) as part of our case, Bill raised several million dollars from foundations not all known for funding such projects (Boeing, Gates, Carnegie and General Electric).



Our NSF grant was for only three years (the most time they were allowed to provide us). The TEDS timeline showed that more time would be required. To provide for later years, Teresa turned out to be very resourceful in getting supplemental grants as well as saving for no-cost extensions. Nevertheless, after grant funds ran low, MSU faculty had to contribute a lot of time for which the university could not be reimbursed.

And until we got the NSF grant in 2005, we had had to rely for three years on a combination of P-TEDS funding and internal MSU resources just to keep us going. Without P-TEDS, we could not have made enough progress to justify the subsequent, larger NSF grant. Moreover, P-TEDS work in six countries (Korea, Taiwan, Bulgaria, Germany, Mexico and the U.S.) was vital to the design of the larger TEDS study. P-TEDS was in effect TEDS on a much smaller scale, with data collected in 2006, making it possible for TEDS to learn from P-TEDS international meetings; from P-TEDS instrument development; from P-TEDS piloting of instruments; from P-TEDS efforts to get cooperation from target institutions, faculty members and students; and from P-TEDS data analyses. We would later be able to claim that many TEDS items were based on five rounds of trialing in multiple countries (initially for P-TEDS and then for TEDS).

Nevertheless, the relationship between P-TEDS and TEDS was never smooth. Bill Schmidt was the first principal investigator for P-TEDS. He was determined that the study be independent and stand on its own with its own publications (Schmidt et al 2007; Schmidt, Blomeke & Tatto, 2011) in addition to developing instruments and methods for the larger TEDS study. He ran the study largely on his own, relying on support from a team of MSU researchers plus a leader from each of the six participating countries. ACER and IEA, sensitive to the possibility of too much MSU influence over TEDS, became more and more reluctant to acknowledge the debt that TEDS owed P-TEDS.

When the P-TEDS and TEDS NSF grants are added to what Schmidt

raised from foundations for U.S. TEDS, the total MSU funding for this whole effort reached a multimillion dollar level rarely achieved by projects in education that are exclusively for research. In contrast, virtually none of the funds from country participation fees (\$32,000 per year per country) was used for MSU. These were needed by ACER, the DPC and the Secretariat who had no other source of funding. In the end, our sole compensation at MSU as we ran out of grant money at the end was our satisfaction in having ourselves raised the grants on which we had drawn.

### *The Hurdles of Organizing for Innovation*

Recruitment of countries for a major new study also required much effort. It was not like TIMSS where many countries had had experience with an earlier version of the same study. Moreover, the IEA organization was built on member institutions which typically were mainly or exclusively concerned with elementary and secondary education, and hence lacked the knowledge, interest and influence in higher education that TEDS required. The following 17 countries ultimately participated in the study: Botswana, Canada (but only four provinces), Chile, Germany, Georgia, Malaysia, Norway, Oman, Philippines, Poland, Russia, Singapore, Spain, German-speaking Switzerland, Taiwan, Thailand and the United States. This was a disappointment because initial expectations were for 30 or more countries and 17 was too few to meet IEA's projected financial breakeven point (based on participation fees and the expenses anticipated at the international level). This preoccupation with breaking even is a good example of the business-like perspective of the new IEA.

Although the number of countries was smaller than desired, within the 17 participating countries an extraordinarily capable group of NRCs emerged. Key participants in P-TEDS Sigrid Bloemeke and Bill Schmidt became the NRCs of Germany and the U.S., respectively. This was a serendipitous arrangement for MSU since otherwise it would have been very difficult to find a suitable role for Bill. In spite of his

extraordinary competence and productivity, I and others felt he did not have the support from other parts of IEA needed for him to be one of the PIs. Nevertheless, I always regarded him as an essential member of the MSU team who could make great contributions to the project. So when Bill himself decided he would be the U.S. National Research Coordinator, that proved to be a great solution to a very ticklish problem.

Another person representing her country in both P-TEDS and TEDS was Feng-Jui Hsieh of Taiwan, an extraordinarily perceptive and thorough researcher in mathematics education at Taiwan National Pedagogical University. She could always be counted upon to find flaws in instruments and documents that no one else was conscious of. Others that were very prominent on the national and/or international scene included Liv Gronmo of the University of Oslo, Norway, and Khoon Yoong Wong of Singapore in mathematics education, Fritz Oser of Switzerland, whose accomplishments were well-known in various areas of educational research, and Beatrice Avalos of Chile, widely published in the general literature on teacher education. Beatrice was particularly amazing in that, although very well known for her writing on teacher and teacher education policies (especially in Latin America but also elsewhere), she was still willing to do so much of the TEDS detail work herself without much assistance.

The challenges and difficulties of TEDS came to a head at periodic international meetings, both the annual meetings of the IEA General Assembly (GA) as well as meetings of the TEDS national research coordinators, expert panel meetings and other specialized meetings. The first GA we attended was in Marrakesh, Morocco, in October 2002. TEDS was the main topic on the agenda for this three-day meeting since it was the only new study being considered at that time. As a result our team had a lot to do before, during and after the meeting. Arriving on Thursday, well before the plenary, we presented and defended our proposal, first to the Technical Executive Group (TEG) on Friday and to the IEA Standing Committee on Saturday. On Sunday

we revised, polished and rehearsed our Powerpoint with heated arguments in preparation for the plenary sessions. Our proposed project was on the agenda several times during the three-day plenary (along with reports on IEA's on-going studies).

Discussions focused primarily on the feasibility of the project. The proposal was, as I see it today, far too ambitious at that point. IEA had originally asked us study teacher education in science as well as mathematics. We also proposed to deal with induction as well as preservice teacher education. And there were advocates of doing more to evaluate the outcomes of teacher education than just test the knowledge of future teachers. No wonder the GA was not immediately convinced that such a sensitive project with these ambitions could be undertaken. But the project was finally approved, subject to certain revisions and follow-up (in this respect we felt like a student in one of our dissertation defenses). We were even required to hold a follow-up meeting in Brussels to allow European representatives to discuss and try to resolve remaining issues before deciding whether to try to get their country to join. Nevertheless, in spite of these conditions that had to be met, at the end of the Morocco meeting TEDS was an official IEA project and no longer just a proposal under consideration.

Ultimately, to respond to feasibility concerns and also because of our own increasing consciousness of what could and could not be done with available time and resources, we had to ditch the science component, due to lack of time and money. It was impossible to multiply the subject-specific work by four to accommodate four science disciplines as well as mathematics (this was the moment when the acronym TEDS became TEDS-M to signify the focus on mathematics). We were also forced to drop the professional development component advocated by our ACER partners, and the collection of data on performance outcomes beyond teacher knowledge.

After the Morocco General Assembly in 2002, the main tasks were to continue to search for funds and to develop what IEA calls a conceptual framework document—a lengthy, published discussion of the

design and rationale for the study (Tatto et al., 2008). This document, part of the new IEA paradigm, reflected a more formalized and public process of design and consultation than I had experienced in previous IEA studies. In provisional form, this document could be used to ensure understanding and consensus on what was to be done and to recruit additional countries. For the purpose of putting this document together, a series of expert panels took place in Amsterdam, East Lansing, and Hamburg. These meetings were partly paid for by IEA and partly by MSU using P-TEDS funds. For the first meeting in Amsterdam in June 2003, the TEDS-M team invited experts in the pure and educational aspects of mathematics and science as well as in teacher education, research methodology and measurement.

All the invited scholars were considered potential participants in further work on TEDS-M. One of our best discoveries was Sigrid Bloemeke who, although a full professor at Humboldt University in Berlin, became known to us only when she was nominated for this meeting. It turned out that she was not only productive and competent in research on teacher education but passionate about it. While participating in P-TEDS and TEDS-M as the national research coordinator for Germany, she came to MSU for two years as a visiting researcher, working mainly with Bill Schmidt's team (2007-2009).

In fall 2003 in Cyprus, we returned to the IEA General Assembly for the second year. We were introduced to the yearly cycle of GA review for all of IEA's current studies, which we then experienced each year until 2010 in Botswana. The first presentation was always to the IEA technical executive group on Friday. Each current study is reviewed twice annually by this panel of internationally well-known methodological experts. The response of the TEG to our Cyprus report was positive – which was not to be the case in later years, and we received a number of compliments on the progress we had made since the previous year's meeting in Morocco. Then the next day we presented to the IEA standing committee, the organization's executive committee.

That also went well. Our main presentation in Cyprus was to the entire General Assembly on Tuesday. Needless to say, by this time, our presentation was becoming quite polished. Even so, we ran into strong criticism from some Europeans who said our study was putting too much emphasis on subject-matter knowledge in the preparation of teachers and not enough on pedagogy and other important aspects of teaching. But in the end, TEDS-M got off relatively easy at this GA, especially when compared not only to our later experience, but also to one of the other projects which had a much rougher time at the Cyprus meeting.

### *My Niches in TEDS-M*

On the teacher education study, the most innovative work I did on instrument development for the teacher education study was to design a questionnaire for those who are primarily responsible for mentoring future teachers during their field experiences, whether these persons be university faculty members, doctoral students, inspectors or other officials from the ministry of education, or the practicing teacher in the classroom to which the future teacher is assigned. Much of the teacher education literature suggests that this mentoring is crucial to the quality of teacher education, and increasingly there is a research literature that deals with mentoring in particular national contexts, including the MSU work in which Sharon Schwille as well as other colleagues have been involved (see e.g. Britton, Paine, Pimm & Raizen, 2003; Dembele, 1995; Schwille, S., 2008; Wang, Odell & Schwille, S., 2008). Finding out more about this role also seemed especially critical simply because many practicing teachers have said that field experience was the only valuable part of their preservice teacher education. I thought this gap was a splendid opportunity for TEDS-M because almost nothing was known about mentoring of field experience on a cross-national basis. However, the Joint Management Committee did not agree and I was overruled in a decision not to use the field experience questionnaire we had developed for the mentors.

They claimed it would be too difficult, almost impossible to collect data from a good probability sample of these mentors. Our ACER colleagues in particular objected. They said that if TEDS-M were to be done in Australia (in fact, it was not), institutions could not collect data from this population because practicing teachers were being designated as mentors but without being closely integrated or held accountable by the university or college. Therefore, the institutions had no complete record of such assignments or even how to identify and get in touch with such mentors. To me, this admission was in itself a telling contrast with countries like Taiwan where the Ministry of Education regulates the field instructor role in detail, documenting an intent to hold the mentors fully accountable for their practice.

Although I continued to participate in meetings and deal with issues of design and instrument development, as the study continued, I spent most of my TEDS-M time on other matters. I did a lot of work on the conceptual framework document, and then gradually shifted my focus to the analysis of policies and context in the participating countries. Originally, we had hoped as in the second civics study to do this analysis before most of the design and instrument development took place so that the results of the policy and contextual analysis would influence the design. But it took so long to get international funding for the study that this sequence did not prove feasible. Instead the investigation of policy and context was carried out more or less simultaneously with work on the surveys of teacher education institutions, faculty members and their students.

The first step in the policy component of TEDS-M was the preparation of country reports. The NRCs were asked to prepare country reports responding to guidelines from our international team. Lawrence Ingvarson of ACER and I took the lead in developing these guidelines. We called for a report that would first discuss context and organization, dealing with historical, cultural and social factors shaping the teacher education system, as well as current policies related to the teacher work force and teacher working conditions, the teacher labor

market, teacher quality issues and finally the structure of the teacher education system. Another part focused more specifically on quality assurance policies in teacher education. When the reports were drafted, Lawrence and I reviewed and edited them, and once we were relatively satisfied, used the reports to write a cross-national synthesis on the issues dealt with in these reports. While on sabbatical at the University of Gothenburg, Sweden, for four months in 2009, I spent much of my time working on this synthesis. Some of this material has also been used in the main international report, and we will have an additional volume with an abridged version of each country report. Through these publications we are attempting to clarify and explain many issues in teacher education from a cross-national perspective, informing scholars, policymakers and practitioners in a way which will be useful in improving policy and practice (Ingvarson & Schwille, forthcoming).

### *Success and Failure*

After the main collection of data, the NRCs of TEDS-M met for the fourth time in Bergen, Norway, in fall of 2008. This meeting served to demonstrate that, whatever happened from then on, TEDS-M was already an extraordinary success. This conclusion was stated most forcefully at the end of the meeting by IEA executive director Hans Wagemaker and by Liv Gronmo, a professor of mathematics education at the University of Oslo and the Norwegian TEDS-M national research coordinator, head of the team that hosted this meeting. They both recalled that, when we began the study, skeptics said it could not be done. These skeptics had asserted, in contrast to TIMSS and other such studies of elementary and secondary education, the institutions, faculty members and students of higher education would never cooperate in a large-scale international survey in which all were asked to fill out fairly lengthy questionnaires, provide syllabi and in the case of the students voluntarily submit to a test of their knowledge of mathematics content and mathematics pedagogy, especially inasmuch as



they were to be selected as part of a national probability sample (or census in the case of the smaller countries) and not because they themselves volunteered. This being the case, skeptics charged, the study would not be able to meet the very demanding IEA technical standards for response rates which are far higher than most surveys of higher education are able to achieve, even ones which do not ask the students to agree to be tested.

As Liv Gronmo, our Norwegian colleague put it, skeptics had the idea that we were building an airplane that was never going to fly. Using this metaphor to great advantage, she asserted that by the time of the Bergen meeting, the plane was flying quite successfully even though she conceded with some understatement that there were bumps here and there. The task that remained was to finish the flight by bringing the plane safely back down to earth. After years of preparation and trial data collection, the final data collection was by then finished and almost all of the countries had submitted data to the IEA data processing center in Hamburg for initial processing, checking and cleaning.

Even with very small countries in which there are very few students in the TEDS-M target populations, the total number of teacher education students for all countries in their last year of teacher preparation who agreed to be surveyed and tested was large – over 15000 students in primary teacher education and 9000 in secondary. In addition, 500 teacher education institutions provided data, as well as almost 5000 faculty members at those institutions. The achieved response rates and coverage of national target populations of students are I believe unmatched in surveys of higher education. Of the 32 national samples of primary or secondary teacher education students, 20 were judged to meet all of IEA's very high standards for response rates, coverage, etc. and were therefore not flagged in the report for any deviations. And even the other samples (with the exception of one country) could be reported with an annotation to show that they deviated but not too much from the IEA standards. Although we might still run into trouble with the data analyses and reporting of all this data, we knew at

that point that there was much credible and unique data that would be of great interest to analyze and report.

IEA had demonstrated that it could do cross national assessment in higher education, overcoming various challenges that exceeded what it had experienced in elementary and secondary education. Moreover, if there were aspects of our forthcoming reports that would not satisfy the critics, they could do their own analyses. IEA policy required that the data it collected be documented, archived and disseminated for secondary analysis by any researcher who wants to do this.

But it must be said that after the 50th anniversary year, setbacks were in store for TEDS-M. In 2009 we were back at the IEA General Assembly meetings in Talinn, Estonia, for our annual report. By this time we had been making this GA trek for seven years since Morocco 2002. According to the earlier timelines, we should have been ready to present our first international report which was to have been made available to GA members shortly before being released to the public in January 2010. But in fact we were far behind. In a rebuke that hit us hard, Hans Wagemaker told us this was the first IEA study to come in late “on his watch.” Even by the 2010 GA in Botswana, we did not have the written report ready to submit; for the GA in Botswana an oral PowerPoint presentation served as the final report to the GA, and as justification for the study to be brought to an official close. The manuscript of the international report was submitted for IEA publication review, publication design and copy editing two months later (Ingvarson & Schwille, forthcoming ; Tatto et al, in press). As of this writing, it still has not been released.

In the meantime, the national centers had been authorized to issue national reports before the international reports - contrary to IEA's normal procedures. Germany, Switzerland, Poland, Taiwan and the U.S. quickly released their reports (e.g., Schmidt et al, 2010; Bloemeke, Kaiser & Lehmann, 2010 a, b--the latter two volumes totaled 800 pages!). This made me feel better about the delays in producing the

international report because I knew that, even so, TEDS-M had begun to be publicly documented and disseminated in the public arena where it could be studied and discussed as a source of evidence and ideas for anyone who was interested.

### **Concluding Reflections**

IEA in 2008 is a research organization which routinely carries out huge, complex surveys of high quality in mathematics, science and reading literacy, using what I would call the Boston College-IEA paradigm. But this paradigm is also restrictive in limiting what is done and applies only partially to new challenges as daunting as TEDS-M.

Nevertheless, when I reflect on the six challenges mentioned at the beginning of this memoir, there is no question but that IEA has come a very long way in developing its capacity for sustained high quality work. In addressing the first challenge of identifying appropriate and manageable domains, the cycle of studies in mathematics, science and reading literacy means that much of this challenge has already been met for these core subjects so that the Boston College team and NRCs can focus on areas that need updating. This is not to say that updating is easy, but rather that focusing the effort permits the team to deal effectively with whatever challenges updating might entail.

In contrast, issues of domain loom larger in TEDS-M and even in civic education although major progress has been made in the experience of three civic education studies (as reflected in the conceptual framework for the third civics study, Schulz et al., 2008). The first study of teacher education is just a first cut at this challenge and now that the design has been used for better or worse, it can be the object of scrutiny and criticism as well as appreciation in preparing for future studies and advancing the state of the art. IEA had to go outside its paradigm to do this study, but now it has the beginnings of a more sustainable capacity and a modified paradigm in studies of higher education.

The second challenge of assembling leadership and support teams will always be a major consideration in planning new studies. At Boston College leadership is already in place for the three core subjects and can be updated by selecting individual staff members when vacancies occur. But even in new studies, this challenge is less daunting now because of IEA's development of sustainable leadership that can be assigned to studies from within the secretariat and DPC.

The third challenge likewise can never be solved completely as long as the source of IEA funding at national and international levels is governments and non-profit foundations. Government funding can never be guaranteed for the future since it is subject to changes of priority as ruling parties and officials change. Foundations may be less volatile but over the decades, if not more quickly, their priorities change, too. But again in the case of the repeat studies, a core of support from the U.S. government to cover the international costs of the repeat studies TIMSS and PIRLS has now been maintained for a considerable time. And at the same time, the proportion of international costs covered by such support has been greatly reduced by IEA's practice for the last ten plus years of charging national centers for participation in individual studies, thus covering an increasing portion of international costs. This in my mind has been the most important organizational change in IEA since the initial framework was put in place in the 1960s.

The resource of time is also a part of the third challenge. Having an optimal amount of time to do research is rarely an option, especially when research is funded by governments. Governments want results soon, usually before one can do all that needs to be done. But here again IEA has made great progress in meeting this challenge by sticking to timelines, thanks in large part to the leadership of Hans Wagemaker and the high value he has put on this consideration. The ability of the Boston College team, working with Secretariat and DPC, to make a four and five year cycle of studies work is a colossal achievement. In contrast, I feel badly that TEDS-M fell behind its timeline even though I know that timelines remain much more tenuous in stud-

ies that have not been done before and where it is difficult to predict the time required to develop a consensus on design and instruments and to do analyses satisfactory to the stakeholders.

For the fourth challenge, building the coalitions necessary to make the studies happen, it is much the same story. The repeat studies of the three core subject matters have built up a network of stakeholders over time, and need not create this network from scratch. In contrast, the new studies have to find individuals, organizations and institutions that have the ability to get resources for the studies in all participating countries and to obtain the cooperation and access necessary to carry out the study. Just to make such entities aware of the study is a challenge, and my longtime collaborator Judith Torney-Purta continues to make great efforts to spread the word about what can be learned from civic education studies. In the new studies, it is always possible that different stakeholders will not come to a consensus. This was true in the second mathematics study where the organizations with a stake in mathematics education in the U.S., whose support was necessary to get funding for the study, were at odds with some participating national centers and others who wanted a study focused on other concerns, one that would be more of a straightforward follow-up of the first mathematics study. The result was the split between countries that did the longitudinal study with specific classroom process and pedagogical content knowledge variables and those that stuck to the cross-sectional design and the types of variables used in other IEA studies.

It is difficult to judge how much capacity IEA has built up to deal with the fifth challenge—changes in context—since these changes are often unpredictable and it is even more unpredictable to know what it will take to deal with them. In my view, this is an area where IEA needs to be networked into various related disciplines whose tools may be needed when such changes in context occur. This seems less likely to happen if as seems likely IEA concentrates its resources on the repeat studies of core subject-matters and positions itself to be responsive more and more to governments in determining the nature of the indi-

cators that are needed. In the second civics study, understanding of the contextual change from the Cold War to the Post-Soviet era was much enhanced by reading the works of political scientists, historians and others who were concentrating on this change.

Showing how the sixth challenge has been met—the challenge of institutionalizing the capacity to meet other challenge—has been the theme of this whole chapter. Going back to the beginning, perhaps I should now ask myself if the late Ben Bloom would himself be convinced that the changes made by IEA to continue its work and to improve the quality of what was done were justified or not. In the absence of Bloom, the reader will have to decide.

The world has changed a great deal since I first sat in a University of Chicago classroom in 1963 and heard C.A. Anderson say that at last we have a measure that can be used across countries in studies of educational outcomes. Due to changes in economies, politics, technology and public attitudes, cooperation between countries has inexorably become both easier and more desirable. The inexorable trend is globalization and what I have experienced in my work with IEA over these many years is one of the most important, most fascinating instances of the globalization of educational research. In my view, the existence, results and, yes, the criticisms of these studies have all contributed greatly to the globalization of public and scholarly discourse on education.

By accident of personality and interests, unencumbered by disciplinary affiliations with mathematics, science or language research, I was always more involved in the more unruly, less disciplined side of IEA. I am glad for this because it exposed me to the more experimental, creative and innovative side of IEA. In fact, the IEA studies I found most interesting and insightful included not only the civic education and teacher education studies I worked on myself, but also IEA's studies of literature, written composition, preprimary schooling, ICT and videos of teaching in the TIMSS video studies.

Recently Judith paid me the great compliment in an email saying that

“if there was a ‘thinking outside the box’ award in IEA, you would certainly be the one to receive it” (email, 10/23/2010, used by permission). I am not sure I deserve such a compliment, but I strongly agree that “thinking outside the box” is to be encouraged in IEA and elsewhere. I believe IEA must continue to find ways to create new modes and forms of international research. For this reason, I have always favored an individual membership division of IEA open to all researchers who want to join, provided that they have a substantial track record of international research in education. But they would not be given a free ride. They should be charged substantial membership dues and expected to contribute their own expertise, advice and ideas to the organization by responding to an annual solicitation as well as participating in meetings and task forces as needed. In short, one of their primary roles would be to become a source of innovation and creativity. Unfortunately, in my view, such a division is very unlikely to get off the ground under present conditions.

In my case, having experienced IEA research, worked with such extraordinary collaborators, learned so much about education and research throughout the world – all this has been its own reward. I have no regrets. Being selected as an Honorary Member of IEA at the 2010 General Assembly in Botswana was a much appreciated and surprising bonus adding to the intrinsic satisfaction I already felt.

## References

- Alexander, R. (2000). *Culture & pedagogy: International comparisons in primary education*. Oxford: Blackwell Publishing.
- Beaton, A. E., Martin, M. O., Mullis, I. V. S., Gonzalez, E. J., Smith, T. A., & Kelly, D.L. (1996). *Science achievement in the middle school years: IEA's third international mathematics and science study (TIMSS)*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Beaton, A. E., Mullis, I. V. S., Martin, M. O., Gonzalez, E. J., Kelly, D. L., & Smith, T. A. (1996). *Mathematics achievement in the middle school years:*

- IEA's third international mathematics and science study (TIMSS)*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Berliner, D. C., & Biddle, B. J. (1995). *The manufactured crisis (myths, fraud, and the attack on America's public schools)*. Reading, MA: Addison-Wesley Publishing.
- Bloemeke, S., Kaiser, G., & Lehmann, R. (2010a). *TEDS-M 2008 – Professionelle kompetenz und lerngelegenheiten angehender primarstufenlehrkräfte im internationalen vergleich*. (Professional competence and learning opportunities of primary school teachers in international perspective). New York: Waxmann Publishing.
- Bloemeke, S., Kaiser, G., Lehmann, R. (2010b). *TEDS-M 2008 – Professionelle kompetenz und lerngelegenheiten angehender mathematiklehrkräfte für die sekundarstufe im internationalen vergleich*. (Professional competence and learning opportunities of secondary mathematics teachers in international perspective). New York: Waxmann Publishing.
- Bos, W., & Lehmann, R. H. (1995). *Reflections on educational achievement: Papers in honour of T. Neville Postlethwaite*. New York: Waxmann Publishing.
- Bradburn, N. M., & Gilford, D. M. (Eds.). (1990). *A framework and principles for international comparative studies in education*. Washington, D.C.: National Academy Press.
- Bradburn, N., Haertel, E., Schwille, J. & Torney-Purta, J. (1991). A rejoinder to "I never promised you first place." *Phi Delta Kappan* 72: 774-777.
- Britton, E., Paine, L., Pimm, D. & Raizen, S. (2003). *Comprehensive teacher induction: Systems for early career learning*. Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Burstein, L. (Ed.), in collaboration with T. J. Cooney, R. A. Garden, E. Kifer, C. C. McKnight, D. Robin, D. F. Robitaille, W. H. Schmidt, J. Schwille, K. J. Travers, & R. G. Wolfe. (1993). *The IEA study of mathematics III: Student growth and classroom processes*. Oxford: Pergamon Press.
- Carroll, J. B. (1975). *The teaching of French as a foreign language in eight*



- countries*. International Studies in Evaluation. Stockholm: Almqvist & Wiksell; New York: John Wiley & Sons.
- Center for Mathematics and Science Education (CMSE), Michigan State University (2010). *Breaking the cycle: An international comparison of U.S. mathematics teacher preparation*. East Lansing, MI: College of Education, Michigan State University.
- Comber, L. C. & Keeves, J. P. (1975). *Science education in nineteen countries*. International Studies in Evaluation. Stockholm: Almqvist & Wiksell; New York: John Wiley & Sons.
- Dembele, M. (1995). *Mentors and mentoring: Frames for action, ways of acting and consequences for novice teachers' learning*. Unpublished doctoral thesis, Michigan State University.
- Elley, W. B. (1994). *The IEA study of reading literacy: Achievement and instruction in thirty-two school systems*. Oxford: Pergamon.
- Epstein, E., & McGinn, N. (Eds.). (2000). *Comparative perspective on the role of education in democratization: Part II, socialization, identity and the politics of control*. Frankfurt am Main: Peter Lang.
- Fagerlind, I., Marklund, S., & Chinapah, V. (1991). *Torsten Husen: An educator*. Studies in Comparative and International Education 20. Stockholm: Institute of International Education, Stockholm University.
- Gonzalez, E. J. & Smith, T. A. (Eds.). (1997). *User guide for the TIMSS international database: Primary and middle school years, 1995 assessment*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Gregory, K. D., & Martin, M. O. (2001). *Technical standards for IEA studies: An annotated bibliography*. Amsterdam: IEA.
- Hansson, G. (1975). *Lasning och litteratur*. Svensk skola i internationell belysning II. Stockholm: Institute for International Education, Stockholm University.
- Husen, T. (1967). *International Study of Achievement in Mathematics: A Comparison of Twelve Countries*. 2 vol., New York: John Wiley & Sons.

- Husen, T. (1983). *Incurable academic: Memoirs of a professor*. Oxford: Pergamon Press.
- Husen, T. et al. (1973). *Natur-Orienterande ammen*. Svensk skola i internationall belysning I. Stockholm: Institute for International Education, Stockholm University.
- International Association for the Evaluation of Educational Achievement (IEA). Editions of 1983, 1985, 1986, 1987, 1993 and 1998. *IEA guidebook: Activities, institutions and people*.
- Ingvarson, L., & Schwille, J. (in collaboration with other series editors M.T. Tatto, S. Senk, R. Peck and G. Rowley). *National policies and regulatory arrangements for the mathematics preparation of future teachers in sixteen countries*. Amsterdam: IEA, forthcoming.
- Inkeles, A. (1977). *The International Evaluation of Educational Achievement*. Proceedings of the National Academy of Education. Vol. 4, pp. 139-200.
- Lewis, E. G. & Massad, C. E. (1975). *The teaching of English as a foreign language in ten countries*. International Studies in Evaluation. Stockholm: Almqvist & Wiksell; New York: John Wiley & Sons.
- Martin, M. O., & Kelly, D. L. (Eds.). (1996). *TIMSS technical report volume 1: Design and development*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Martin, M. O., & Kelly, D. L. (Eds.). (1997). *TIMSS technical report volume 2: Implementation and analysis, primary and middle school years*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Martin, M. O., & Mullis, I. V. S. (Eds.). (1996). *TIMSS quality assurance in data collection*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Martin, M. O., Mullis, I. V. S., Beaton, A. E., Gonzalez, E. J., Smith, T. A., & Kelly, D. L. (1997). *Science achievement in the primary school years: IEA's third international mathematics and science study*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Martin, M. O., Mullis, I. V. S., & Foy, P. (2008). *TIMSS 2007 international*

- science report*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Martin, M.O., Mullis, I. V. S., & Kennedy, A. M. (2007). *PIRLS 2006 technical report*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Martin, M. O., Rust, K., & Adams, R. J. (1999). *Technical standards for IEA studies*. Amsterdam: IEA.
- McKnight, C. C., Crosswhite, F. J., Dossey, J. A., Kifer, E., Swafford, J. O., Travers, K. J., & Cooney, T. J. (1987). *The underachieving curriculum: Assessing U.S. school mathematics from an international perspective*. Champaign, IL: Stipes Publishing.
- Mullis, I. V. S., Martin, M. O., & Foy, P. (2008). *TIMSS 2007 international mathematics report*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Mullis, I. V. S., Martin, M. O., Kennedy, A. M., & Foy, P. (2007). *PIRLS 2006 international report*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Mullis, I.V.S., Martin, M. O., Beaton, A.E., Gonzalez, E.J., Kelly, D.L., & Smith, T.A. (1997). *Mathematics achievement in the primary school years: IEA's third international mathematics and science study (TIMSS)*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Munck, I. M. (1979). *Model building in comparative education: Applications of the LISREL method to cross-national survey data*. IEA Monograph Series No. 10. Published doctoral thesis, University of Stockholm.
- Muthen, B. (1991). Analysis of longitudinal data using latent variable models with varying parameters. In L. Collins & J. Horn (Eds.) *Best Methods for Analysis of Change*, (pp 1-17). Washington, DC: American Psychological Association.
- Nielsen, H. D. (1977). *Tolerating political dissent*. IEA Monograph Studies No. 6. Published doctoral thesis, University of Stockholm. Almqvist & Wiksell.

- Passow, A. H., Noah, H. J., Eckstein, M. A., Mallea, J. (1976). *The national case study: An empirical comparative study of twenty-one educational systems*. International Studies in Evaluation. Stockholm: Almqvist & Wiksell; New York: John Wiley & Sons.
- Peaker, G. F. (1975). *An empirical study of education in twenty-one countries: A technical report*. International Studies in Evaluation. Stockholm: Almqvist & Wiksell; New York: John Wiley & Sons.
- Porter, A. C. & Gamoran, A. (Eds.). (2002). *Methodological advances in cross-national surveys of educational achievement*. Washington DC: National Academy Press.
- Postlethwaite, N. (1967). *School organization and student achievement: A study based on achievement in mathematics in twelve countries*. Stockholm Studies in Educational Psychology 15. Published doctoral thesis, University of Stockholm.
- Purves, A. (1975). *Literature education in ten countries*. International Studies in Evaluation. Stockholm: Almqvist & Wiksell, New York: John Wiley & Sons.
- Purves, A. C. & Levine, D. M. (1975). *Educational policy and international assessment: Implications of the IEA surveys of achievement*. Berkeley: McCutchan.
- Robitaille, D. F., & Garden, R. A. (Eds.). (1989). *The IEA study of mathematics II: Contexts and outcomes of school mathematics*. Oxford: Pergamon Press.
- Schleicher, A. (1994). *Standards for the design and operations in IEA studies*. Paper presented at the Seminar on Achievement in School and Attitudes towards the Key Questions about the Future, sponsored by Statistics Sweden, Stockholm, June 3, 1994.
- Schmidt, W. H. et al. (1996). *Characterizing pedagogical flow: An investigation of mathematics and science teaching in six countries*. Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Schmidt, W. H., McKnight, C. C., Cogan, L., Jakwerth, P., & Houang, R. (1999). *Facing the consequences*. Dordrecht, The Netherlands: Kluwer

Academic Publishers.

- Schmidt, W. H., Blomeke, I., Tatto, M. T. (2011). *Teacher education matters: A study of middle school mathematics teacher preparation in six countries*. New York: Teachers College Press.
- Schmidt, W. H., McKnight, C. C. & Raizen, S. (1997). *A splintered vision: An investigation of U.S. science and mathematics education*. Dordrecht, The Netherlands: Kluwer Academic Press.
- Schmidt, W. H. et al. (2001). *Why schools matter: A cross-national comparison of curriculum and learning*. San Francisco: Jossey-Bass.
- Schmidt, W. H., et al. (2007). *The preparation gap: Teacher education for middle school mathematics in six countries (MT21 Report)*. East Lansing, MI: College of Education, Michigan State University.
- Schulz, W., Fraillon, J., Ainley, J., Losito, B. & Kerr, D. (2008). *International civic and citizenship study: Assessment framework*. Amsterdam: IEA.
- Schwille, J. (1975). Predictors of between student differences in civic education cognitive achievement. In J. Torney et al., *Civic Education in Ten Countries*. Stockholm: Almqvist & Wiksell; New York: John Wiley & Sons, Halsted Press.
- Schwille, J. (1997). *From over-emphasis on an official curriculum to analysis of multiple perspectives on preparation for citizenship: Advances in the design and implementation of an IEA study*. Paper presented at annual meetings of the American Educational Research Association, Chicago, IL.
- Schwille, J. (2010). *Torsten Husen and Neville Postlethwaite: Two intertwined and extraordinary careers*. Paper prepared for memorial session at annual meeting of the Comparative and International Education Society, Chicago, IL.
- Schwille, J. et al. (1975). *IEA six subject survey data bank manuals*. Stockholm: IEA.
- Schwille, J., & Amadeo, J. (2002). The paradoxical situation of civic education in school: ubiquitous and yet elusive. In G. Steiner-Khamsi, J. Torney-Purta & J. Schwille (Eds.). *New Paradigms and Recurring*

- Paradoxes in Education for Citizenship: An International Comparison.* Amsterdam: Elsevier.
- Schwille, J., & Burstein, L. (1987). The necessity of tradeoffs and coalition building in cross national research: a critique of Theisen, Achola and Boakari. *Comparative Education Review* 31, 602-611.
- Schwille, J., & Dembele, M. (2007). *Global perspectives on teacher learning: Improving policy and practice.* Fundamentals of Educational Planning Series. Paris: International Institute for Educational Planning (IIEP).
- Schwille, S. (2008). The professional practice of mentoring. *American Journal of Education.* 115: 139-167.
- Steiner-Khamsi, G., Torney-Purta, J., & Schwille, J. (Eds.). (2002). *New paradigms and recurring paradoxes in education for citizenship: An international comparison.* Amsterdam: Elsevier.
- Super, D. E. (Ed.). (1967). *Toward a cross-national model of education achievement in a national economy: The report of the Lake Mohonk conference.* New York: Columbia University Teachers College.
- Tatto, M. T., Nielsen, H. D., Cummings, W. C., Kularatna, N. G., & Dharmadasa, D.H. (1993). Comparing the effectiveness and costs of different approaches for educating primary school teachers in Sri Lanka. *Teaching and Teacher Education*, 9(1), 41-64.
- Tatto, M. T., Schwille, J., Senk, S.L., Bankov, K., Rodriguez, M., Reckase, M., Ingvarson, L., Rowley, G., & Peck, R. (in press). *The teacher education study in mathematics (TEDS-M), international report: Findings from the IEA study of the mathematics preparation of future teachers.* Amsterdam: IEA
- Tatto, M. T., Schwille, J., Senk, S., Ingvarson, L., Peck, R. & Rowley, G. (2008). *Teacher education and development study in mathematics (TEDS-M): Policy, practice, and readiness to teach primary and secondary mathematics--conceptual framework.* Amsterdam: IEA.
- Theisen, G. L., Achola, P. P. W., & Boakari, F.M. (1983). The underachievement of cross-national studies of achievement. *Compa-*

- rative Education Review* 27 (1), 46-68.
- Thorndyke, R. L. (1975). *Reading comprehension education in fifteen countries*. International Studies in Evaluation. Stockholm: Almqvist & Wiksell; New York, John Wiley & Sons.
- Torney, J. V., Oppenheim, A.N., & Farnen, R. F. (1975). *Civic education in ten countries*. International Studies in Evaluation. Stockholm: Almqvist & Wiksell; New York, John Wiley & Sons.
- Torney Purta, J., & Schwille, J. (1986). Civic values learned in school: policy and practice in industrialized countries. *Comparative Education Review* 30, 30-49.
- Torney-Purta, J., & Schwille, J. (2011). The IEA studies of civic education, 1970-2005. In C. Papanastasiou. & T. Plomp (Eds.) *IEA 1958-2008 - 50 years of experiences and memories*. Nicosia: Kailas
- Torney-Purta, J., Amadeo, J. & Schwille, J. (2003). Political democracy and the IEA study of civic education. In J. Guthrie (Ed.), *Encyclopedia of Education* 2nd ed., (pp. 1236-1238) NY: Macmillan Reference.
- Torney-Purta, J., Amadeo, J. & Schwille, J. (2010). International comparisons: the IEA study of civic education. In E. Baker, B. McGaw & P. Peterson (Eds.), *International Encyclopedia of Education* 3rd ed. Amsterdam: Elsevier.
- Torney-Purta, J., Lehmann, R., Oswald, H. & Schulz, W. (2001). *Citizenship and education in twenty-eight countries*. Amsterdam: IEA.
- Torney-Purta, J., Schwille, J. & Amadeo, J. (Eds.). (1999). *Civic education across countries: Twenty-four national case studies from the IEA civic education project*. Amsterdam: IEA.
- U.S. Department of Education, National Commission on Excellence in Education. (1983). *A nation at risk, the imperative for education reform: A report to the nation and the secretary of education*. Washington DC: U.S. Department of Education.
- Valverde, G. A., Bianchi, L. J., Wolfe, R. G., Schmidt, W. H., & Houang, R. T. (2002). *According to the book: Using TIMSS to investigate the*

*translation of policy into practice through the world of textbooks.* Dordrecht: Kluwer.

Walker, D. A. (1976). *The IEA six subject survey: An empirical study of education in twenty-one countries.* International Studies in Educational Evaluation. Stockholm: Amquist & Wiksell; New York: John Wiley & Sons.

Wang, J., Odell, S. J. & Schwille, S. A. (2008). Effects of teacher induction on beginning teachers' teaching: A critical review of the literature. *Journal of Teacher Education* 59 (2), 132-152.

Wang, Q. (2010). *Matching for bias reduction in treatment effect estimation of hierarchically structured synthetic cohort design data.* Unpublished Ph.D. dissertation, Michigan State University.

Westbury, I., & Travers, K. (1990). *Second international mathematics study: An overview of IEA study.* Urbana, Illinois: College of Education, University of Illinois.

Wolf, R. M. (1977). *Achievement in America: National report of the United States for the international educational achievement project.* New York: Teachers College Press.







**PART E:**  
**MEMORIES OF FRIENDS OF THE IEA**



## CHAPTER 28

# Reflections on IEA from the Perspective of a World Bank Official

*Marlaine E. Lockheed*

Over the nearly four decades that I have known the IEA I have seen it grow in size, diversity and technical strength. It has been an honor to be associated with this organization, in even the modest way that I was involved, while at the World Bank and subsequently after retirement. My first encounter with IEA was not in the context of the World Bank, however, but a decade before, as a research scientist at Educational Testing Service (ETS), where – as a sociologist – I acquired a deep respect for psychometricians, statisticians and the technical complexities of standardized testing. Sent by ETS to an International Symposium on Educational Testing, held in The Hague in April 1974, I encountered there many of the distinguished scholars who were involved in the first IEA studies, such as Neville Postlewaithe, Leigh Burstein and Richard Shavelson (both contemporaries of mine from graduate school at Stanford), as well as others I would later come to admire for their constant and vigorous support for the concept of international assessments.

While at the World Bank and since retirement, I have had three types of experiences with IEA. In the 1980s, I was a research user of IEA data sets in the Education Department of the Bank; in the 1990s I monitored Bank grants to IEA and was the Bank's official "observer" at various IEA General Assembly (GA) meetings; and recently – since retirement – I served as an independent evaluator of the most recent series of World Bank grants to IEA. This has given me a broad view of how IEA has changed and how it has influenced education research and policy worldwide.

## IEA Data for World Bank Policy

IEA data played a key role in the World Bank's education policy work in the 1980s and 1990s, both in describing the level of achievement of students in developing countries and in exploring factors associated with higher achievement. IEA data are highlighted in two major documents from that time, *Primary Education: A World Bank Policy Paper* (1990) and *Priorities and Strategies for Education* (1995); IEA is also mentioned as a Bank "partner" in the 1999 *Education Sector Strategy*.

In the late 1980s, the World Bank's Human Development Department prepared the Bank's first-ever policy paper on basic education. It was evidence-based, intended to provide guidance to countries about cost-effective investments in educational inputs, and summarized what was then known about education in developing countries and "what worked" to boost learning. As the paper's chief author, I was responsible for creating and assembling this evidence.

At that time, evidence regarding the levels of learning in developing countries was scant, with cross-national data almost exclusively derived from the earliest IEA studies of reading, math and science achievement; little was also known about what improved student learning in the developing world. While any number of studies had looked at correlates of higher test scores in developed countries, only a very few studies had looked at such factors in low and middle-income countries. Three World Bank papers— the first by John Simmons (Simmons & Alexander, 1974), the second by Steve Heyneman and William Loxley (Heyneman & Loxley, 1983), later an Executive Director of IEA, and the third by Bruce Fuller (Fuller, 1987)—summarized the extant literature, which drew extensively on the earliest IEA studies.

But by the mid-1980s, the second international mathematics and science studies had been completed, providing a new opportunity to examine learning achievement in developing countries. I contacted Malcolm Rossier, the second international science study coordinator, and Nongnutch Wattanawaha, coordinator of the second mathematics

study for Thailand, both of whom kindly provided the Bank with the data for secondary analysis, enabling us to examine student learning achievement in four developing countries – Nigeria, the Philippines, Swaziland and Thailand. In addition, Eduardo Luna provided us data from a study in the Dominican Republic that replicated the design and instruments used IEA's second mathematics study. With these data sets, we were able to answer specific questions of concern to the Bank: Were private schools more effective or efficient than public ones? What inputs were associated with higher learning achievement? How important are the qualifications and classroom practices of teachers? Does local control enhance school effectiveness? What were the effects of single-sex schools?

As a result of this cooperation, the Bank's policy paper on primary education not only included summary data from the first rounds of IEA studies, it also presented the findings from this new program of research (a list of these studies is attached). The Bank policy paper was accompanied by a book, *Improving Primary Education in Developing Countries*, which highlighted IEA's cross-national studies of learning, served as the World Bank's contribution to the Global Conference on Education for All, in Jomtien, Thailand in 1990, and contributed to the Jomtien Declaration's strong emphasis on the importance of measuring students' learning (Lockheed & Vespoor, 1990). Shortly thereafter, the *International Journal of Educational Research* published a special issue devoted to research on private versus public education; it featured several articles based on IEA data and from this program of research (Jimenez & Lockheed, 1991).

IEA studies were also highlighted in the 1995 Bank sector paper *Priorities and Strategies for Education*, which drew on the reading literacy study of the early 1990s to show the sharp differences in reading achievement between developing and developed countries, to document the wide variation among schools in developing countries with respect to average student achievement, and to underscore the huge rural-urban gap in reading. This analysis was prepared for the Bank by

IEA's then director of data management and analysis, Andreas Schleicher.

In part because of the Bank's long-standing relationship with IEA, another Bank activity, EdStats, expanded to include IEA data. EdStats is a partnership between the Bank, UNESCO Institute of Statistics and IEA; it aims to disseminate education data inside and outside the Bank and provides links to a large number of education websites worldwide. Data on student learning achievement from TIMSS, PIRLS, PISA, LLECE, PASEC and SACMEQ are reported.

### **World Bank Support for IEA Studies**

World Bank support for cross-national studies of learning focused both on creating country-level demand for information about student achievement and on providing resources for generating it, responding to the dearth of countries for which such data were available. Even though World Bank education policy documents in the 1990s referenced student achievement, the reality was that as of 1990 only a handful of developing countries had participated in the early IEA studies. In the 1970s, 4 low and middle income countries participated in the first reading study, 5 in the first science study, and 5 in the second math study; in the 1980s, 11 developing countries participated in IEA's second science study and 10 in the reading literacy study. Even though this increase was substantial, cross-national measures of learning achievement were unavailable for 90 percent of low and lower-middle income countries. By comparison, data on school enrollments were available for nearly 100 low and middle-income countries. Something needed to be done to encourage more developing countries to generate cross-nationally comparable data on learning, and the World Bank stepped up to this challenge, in three ways.

The first – demand-generation – approach of the early 1990s was to provide training about testing and assessment to World Bank staff and

their country counterparts. Bank staff in the education sector participated in two week-long training courses offered by Educational Testing Service (ETS) in conjunction with the Bank's Education and Employment Division, while country teams participated in various regional courses on assessment offered by the World Bank Institute (then called the Economic Development Institute). These courses were intended to raise staff and counterpart awareness about testing issues and create a demand to include a "testing component" in Bank-financed education projects. To assist countries prepare such projects, the resource materials provided to participants included the resumes and contact information for testing experts world-wide, including such IEA study directors as Alan Dock, John Izard, Michael Martin, Malcolm Rosier, John Schwille, Yahya Umar - some of whom later became World Bank consultants on testing components of projects. My involvement was to conceptualize and implement the staff training and to serve as an "expert" for some of the country training.

Once demand had been created within the Bank and at the country level, the second approach was for World Bank-funded education projects to include a small "testing component" financed by loans or credits; the "testing component" could strengthen local capacity for testing and assessment as well as enable countries' participation in international assessments. These testing components were very small, and for them to be included in larger education sector projects required considerable personal involvement from both concerned Bank staff (task managers and division managers) and from the IEA study coordinators or potential coordinators in the country. But progress was rapid, and by the early 1990s, 50 percent of Bank education projects supported a "testing component", up from 5 percent in 1975, and since then well over 60 percent of projects have included such a component. Some of these projects found local support from earlier IEA study coordinators, such as Shukla Sharma in India and Josefina Fonacier in the Philippines. Not all testing components of World Bank-financed projects have supported international studies, of course, but the



growth in the number of developing countries participating in IEA studies, beginning in the early 1990s, suggests that this country-level support may have been a factor.

The third World Bank approach was to fund the IEA directly, through grants that would enable low and lower-middle income countries to participate in the international activities, particularly training, associated with an IEA study. This approach resulted in a dramatic increase in the number of developing countries participating in IEA studies. In the late 1990s, the first Bank grant helped 18 low- and middle income countries<sup>1</sup> participate in TIMSS-1999 (also known as TIMSS-R). The actual amount of support was quite small, about 11% of the total costs of the IEA study, and typically covered the costs of the international fee and international travel to IEA meetings and training; in a few cases the Bank also covered some direct out-of-pockets costs involved in carrying out the studies. In the early 2000s, the Bank again grant-funded IEA so that 15 low- and middle-income countries could participate in PIRLS and 21 low- and middle-income countries could participate in TIMSS-2003. And finally, in 2005, the Bank provided grant support to the Program for the Assessment of Student Achievement (PASA), an IEA-led consortium of regional cross-national assessment and policy bodies,<sup>2</sup> with a broad mandate related to improving assessment capacity in low- and middle-income countries. As a consequence of these grants, the number of developing countries participating in TIMSS-1999, TIMSS-2003, TIMSS-2007, PIRLS-2001 and PIRLS-2006 soared.

It was in the context of initiating and managing these support activities that I became more closely involved in IEA, including attending

---

1. Bulgaria, Hungary, Indonesia, Jordan, Latvia, Lithuania, Macedonia, Malaysia, Moldova, Morocco, Philippines, Romania, Slovak Republic, Slovenia, South Africa, Thailand, Tunisia and Turkey.

2. The other participants were the Partnership for Educational Revitalization in the Americas (PREAL), Southern Africa Consortium for the Measurement of Education Quality (SACMEQ), Programme d'Analyse des Systems Educatifs de la Confemen (PASEC) and the West African Examinations Council (WAEC).

General Assembly meetings as an official World Bank observer off and on throughout the 1990s. These meetings confirmed that the IEA was largely a membership organization of men from developed countries (with a few notable exceptions), and that the governance of IEA as well as the technical and managerial aspects of the studies were also largely dominated by distinguished male scholars from North America and Europe, again with a few notable exceptions such as Judith Torney-Purta, the first woman to be accorded an honorary member of IEA. A quick scan of the ten men who were individual and honorary members of IEA in the late 1990s confirms this perception about the early leadership of the organization : Zoltan Bathory of Hungary, Benjamin Bloom of the USA, Robert Garden of New Zealand, John Keeves of Australia, Gilbert de Landsheere of Belgium, Torsten Husen of Sweden, Roy Phillipps of New Zealand, Neville Postlethwaite of the UK, Alan Purves of the USA, Richard Wolfe of the USA, and Aldo Visalberghi of Italy. No women, and only one from a low or middle-income country.

During this period, I also I served as the World Bank's representative on the National Academy of Science Board on International and Comparative Studies in Education (BICSE), chaired by Gordon Ambach, a US representative to the IEA. BICSE established to provide guidance to the US government with regard to studies worthy of support by the National Science Foundation and the National Center for Education Statistics. As a result, one regular agenda item for the quarterly BICSE meetings was a briefing on the progress of IEA studies by the international coordination center directors, Al Beaton and Ina Mullis. BICSE also sponsored workshops and conferences around themes that could be informed by international studies; for example, a workshop held in 1996 highlighted the role of TIMSS in establishing international standards for mathematics and science (Beatty, 1997). These BICSE meetings revealed the strong support for the international assessments that Larry Suter, of the US National Science Foundation, and Eugene Owen and Lois Peak, of the US National Center for Education Statistics, had created within their own organi-

zations, which helped solidify support within the Bank, since a condition of the grant program was that other donors participate in funding the activity. Thus, although World Bank grants did not fund countries participating in the third round of mathematics and science studies (TIMSS-1995), they did play a crucial role in financing low- and middle-income countries wishing to participate in subsequent TIMSS and PIRLS studies.

### **Evaluating IEA's Influence**

My third association with IEA followed my retirement from the Bank in 2004. IEA asked me to carry out an independent evaluation of its most recent grant from the World Bank, given to a consortium of institutions collectively known as the Program on the Assessment of Student Achievement (PASA). The purpose of PASA was to enhance the capacity of IEA for comparative international assessment and region-specific support, and to enhance the capacity of regional assessment institutions in Africa and Latin America for comparative regional and sub-regional educational assessment and the use of assessment information.

In conducting this evaluation, I once again served as an observer at a GA meeting (in Hong Kong in 2007) and was struck by the dramatic change in personae. Dozens of representatives from low and middle-income countries were present, reflecting the sea change in the membership of IEA, in which now over a third of the official members of IEA are education systems of developing countries. In addition, I attended a meeting of the national coordinators for TIMSS 2007 and saw the same change, with more developing countries than developed countries participating in the study (58 percent) and a group of coordinators that seemed evenly divided between men and women. Moreover, the managerial and technical committees of the IEA had also changed dramatically, to include more experts from developing countries.

One of the conclusions of my independent evaluation was that a country's participation in an IEA study appears to strongly build the coun-

try's capacity for assessment and participation in subsequent studies, even under different sponsorship. For example 76 percent of the developing countries participating in the first OECD study, PISA 2000, and 68 percent of the developing countries participating in PISA 2006 had previously participated in at least one IEA study. Participation in international assessments also appears to be related to regularity in national assessments, with 41 percent of developing countries that have carried out at least one "national assessment" (including census-based and high stakes assessments) have participated in at least one IEA study.

In part, this may be due to the increased attention that IEA has paid to training country teams from developing countries. A case in point is the series of training sessions provided to national coordinators and technical specialists from countries in the Middle East and North Africa, under funding provided by the World Bank to PASA. In these seminars, participants carried out secondary analyses of earlier PIRLS and TIMSS data, and were helped to write policy documents based on these analyses. In addition, throughout the cycle of conducting an IEA study, country coordinators receive significant "hand-on" training related to developing the tests, sampling, managing data collection, scoring tests and preparing reports, which was highly valued by participants.

Other results from my independent evaluation included the influence of the IEA studies on educational policy in the participating countries, on such international documents as the UNESCO Global Monitoring Reports, and on more recent World Bank documents, such as the MENA region's "flagship" report, *The Road Not Travelled: Education Reform in the Middle East and North Africa* (2008) that has a chapter reporting regional results on TIMSS 2003. IEA studies have also contributed significantly to the scholarly literature related to education quality, with TIMSS and PIRLS the most visible studies. My evaluation also found that direct training and participation in various IEA activities has contributed to improved capacity for assessments, as indicated by participant self-reports, regularity of participation in international assessments and regularity of national assessments. Interviews

held in December 2009 with TIMSS 2011 national coordinators underscored the importance of the IEA studies in framing curriculum reform in their countries.

In short, therefore, IEA has transformed itself from a small society of Western male educators and psychometricians to a global institution with representation from virtually all continents. Its influence has spread to other international and regional assessments through a network of scholars who worked on early IEA studies. And IEA studies have become a means of building a network of experts working on assessments around the globe.

## References

- Beatty, A. (Ed.). (1997). *Taking stock: What have we learned about making education standards internationally competitive?* Washington DC: National Research Council.
- Fuller, B. (1987). Raising school quality in developing countries: What investments boost learning? *Review of Educational Research*, 57, 255-292.
- Heyneman, S., & Loxley, W. (1983). The effect of primary school quality on academic achievement across twenty-nine high and low income countries. *The American journal of sociology*, 88 (6), 1162-94.
- Jimenez, E. & Lockheed, M. (Eds). (1991). Private versus Public Education: An International Perspective. Special Issue. *International Journal of Educational Research*, 5.
- Lockheed, M. (2010). *The craft of education assessment: Does participating in international and regional assessments build assessment capacity in developing countries?* Report submitted to IEA and the World Bank.
- Lockheed, M., & Verspoor, A. (1990). *Improving Primary Education in Developing Countries*. Oxford University Press.
- Simmons, J., & Alexander, L. (1974). *The determinants of school achievement: Educational production function analysis*. Washington DC: The World Bank.

## CHAPTER 29

# **Guiding International Comparative Studies (1981 to 2009): Reflections of a U.S. Program Director**

*Larry E. Suter*

The International Association for the Evaluation of Educational Achievement (IEA) provided me - a government statistician - with the opportunity to discover new ways of defining, describing, and explaining student learning through seeking similarities and differences in educational practices between different countries. In this chapter I will respond to a request from Tjeerd Plomp and Constantinos Papanastasiou to reflect on my participation in the IEA between 1980 and 2010. During that 30 year period, the IEA grew from a loose federation of academic researchers from a few countries, along with some supportive and interested employees of government agencies, to a large survey organization that introduced the methods of measurement of student achievement to many new countries. The leaders of the IEA grew out of the experiences of a few internationally known experimental educational psychologists who in the 1950's had conceived of conducting cross national comparisons: amongst others Benjamin Bloom and Robert Anderson of the University of Chicago (Anderson was the recipient of the first grant from the U.S. Office of Education), Torsten Husén of the University of Sweden and Robert Thorndike of Columbia University (see the chapter in this volume by John Keeves for additional information).

My own involvement with the IEA began officially in 1981 when the National Center for Education Statistics (NCES) established a contract to create an international data base. Later on at the National Science Foundation from 1990 to 2010, I managed the funding of several research and development proposals that involved the IEA. Altogether these projects included the Second International

Mathematics Study, the Second International Science Study, the International Reading Literacy Study, the International Pre-School Study, The Computer Education study (COMPED), and the International Assessment of Education Progress (IAEP), Third International Mathematics and Science Study (TIMSS), the TIMSS video study, and the Teacher Education Study (TEDS).

This chapter reviews the decision making process in which the U.S. government funded international education comparative studies from 1981 to 2010. It notes the roles of specific decision makers and advisors to the government funding agencies and the contributions of many researchers who conducted the major studies in this period. The chapter comments on the contributions that the IEA studies have made to policy discussions of educational practices especially in mathematics and science. Many more significant actors were involved in these studies than are mentioned here. Those discussed are believed to be “at the cutting edge” of the decision making processes that resulted in the growth and acknowledgement of international comparative education studies.

### **The Second Round of International Studies: The Re-making of an International Organization: 1981 to 1989**

In 1981 Marie Eldridge, the Commissioner of NCES, recognized a need in the organization to improve the availability of statistics of the United States compared with other countries. The comparative statistical resources at OECD and UNESCO were known to NCES because the agency provided data to them for the United States. Also, the agency was responsible for producing the surveys for the National Assessment of Education Progress (NAEP) which was the main source of U.S. national survey data on student achievement. But in 1981, no one at the agency was familiar with organizations that conducted student assessments internationally. Steven Heyneman of the World Bank informed us that the IEA had just initiated a second round of studies

concentrating on mathematics learning called the Second International Mathematics Study and that the U.S. portion had been partially funded by the National Science Foundation. While the leaders of this particular study were in New Zealand (Roy Phillips), an American, Dr. Kenneth Travers of the University of Illinois, was the chair of the International Mathematics committee that was preparing the framework for the study of mathematics and was easily accessible to a U.S. government agency.

In March 1981, Dr. Travers agreed to discuss how to create an international data base from the Second International Mathematics Study (SIMS) that might provide NCES with the first comparative data on student achievement. Unfortunately, the field work for the U.S. portion of the international survey had already begun so NCES had no opportunity to affect the field work for the U.S. data collection. The previous work of the IEA was known by Susan Chipman, a mathematics education specialist in the National Institute of Education (also a part of the Office of Education), who encouraged us to develop a contract with the University of Illinois to create an international database that could be used by our agency in its summary publications and by national researchers. Through special circumstances, the agency had resources available that year to support the costs of a contract with the Dr. Travers at the University of Illinois to support the activities of the U.S. mathematics committee for purposes of writing papers about the data collection procedures, opportunity to learn, and student achievement, and to create an international data base for others to use. The international center for the mathematics study was located in New Zealand and the leader of the IEA, Neville Postlethwaite, was in Germany, while the funding for the international database was going to Illinois as a contract with specific objectives.

An agreement was reached between the IEA international leaders and NCES during a meeting in Toronto in 1983. At that meeting Torsten Husén, who had recently participated in the US Commission on



Excellence, Neville Postlethwaite, Richard Wolf, Roy Phillips, and John Keeves discussed NCES collaboration with the upcoming international mathematics study. These leaders were happy to have support for their international efforts but they were wary of this new U.S. interest in their study especially since it required a U.S. government contract with the U.S. mathematics committee to produce specific products (a database and written reports). An agreement was reached to involve the international organization in major decisions regarding the use and release of data gathered in other countries. In many ways, the contract with the University of Illinois may have been responsible for the continued survival of the IEA itself because it provided financial support during a critical period for the continued development of the international data base, for meetings, and for analyses that were to come later.

The framework for the data collection and analysis of the Second International Mathematics Study was unlike any previous US study of student achievement because it focused on the educational practices of teachers and as well as on the performance of students. The US mathematics committee chaired by Kenneth Travers (Illinois), contained others who had significant impact on the study: Joseph Crosswhite (president of the National Council for Teachers of Mathematics), John Dossey (Mathematics education, Illinois State), William Schmidt (Statistics, Michigan State), Richard Wolfe (of the Ontario Institute for Education, Toronto), Skip Kifer (U. of Kentucky), and Leigh Burstein (UCLA). These investigators believed that the focus of this study should be on the content of the curriculum available to students and they experimented with new methods for measuring how the Opportunity to Learn mathematics could be traced to the classroom practices and school policies. The curriculum was conceived as being “intended” at the national level; “implemented” by classroom teachers, and “attained” by students. To measure each of these levels of expectation in mathematics, the SIMS created measurement instruments for the schools, teachers, and students that followed a theme of

particular aspects of mathematics so that the final statistics could illustrate whether a topic in mathematics was supposed to be taught, whether it was actually taught, and whether the students learned it or not. A report by Robert Garden (New Zealand department of education) from SIMS (Garden, 1982) summarized the student level performance and the classroom coverage in a clear manner for all participating countries. This concept (although not original with this study) grew to become a widely adopted framework for conducting research on student learning. While the international studies were being written and published, the American National Council of Teachers of Mathematics were preparing their first publication on Standards for Mathematics. The authors of that report were also engaged with SIMS (especially Joe Crosswhite and John Dossey). Thus, the design and implementation of the IEA study of mathematics had an influence on the development of national standards in mathematics. The NCTM report of 1989 became a model for other fields and the effort to define standards continues into the 21st century.

Leigh Burstein had been a student of Lee Cronbach at Stanford and had spent a year in Stockholm as a graduate student fellow during the early 1970's working with the data from the first round of IEA studies. Leigh was a loud spokesman for the IEA during the period of the SIMS. He personally wrote letters, made visits, and phoned the head of NCES (Emerson Elliott) about the importance of supporting basic research on assessment and learning as being carried out in institutions of higher education. His passion for the IEA and his belief in the value of extensive analysis of the curriculum and teaching practices was responsible at least in part for the acceptance of the IEA studies by the Federal Agency NCES. He edited and wrote the third volume of the IEA Second International Mathematics Study that examined longitudinal measures of the final year of secondary school (Burstein 1989). Leigh died of a heart attack during a meeting of the US advisory committee for TIMSS in 1993 while running for exercise in the morning of the meeting to review the US plans for analysis. In many ways, Leigh

Burstein is responsible for the US involvement in the IEA because he believed that the IEA method of involving researchers rather than study administrators would advance methodology and knowledge of the assessment of student performance in educational systems.

### **How the International Study Results Became Important to Policy**

By chance, policy makers at very high levels in the United States became very interested in the results of international studies following the 1983 US Commission on Excellence report of the status of US education, "A Nation at Risk". The commissioners were interested in obtaining international comparative data for their deliberations and asked Torsten Husén to meet with the staff to provide an international dimension to their deliberations. However, the most significant event was the discovery of a paper that had been published by Barbara Lerner a lawyer who had been a student of Benjamin Bloom at the University of Chicago (Lerner, 1982; Wolf 1977; 1983). Lerner had re-analyzed the results of the first IEA studies to create "league tables" that were not produced in the first IEA reports and concluded that the results indicated that the growth in U.S. investment in education had not been effective because our standing compared with other countries was low. This paper provided the context that the leaders of the Commission were seeking. They felt that international comparisons was the most powerful way to make the point that schools in the United States needed the attention of the President himself. The phrases used in that report by are still quoted today. The introductory paragraph to the Commission report, "A Nation at Risk" says, ". . . the educational foundations of our society are presently being eroded by a rising tide of mediocrity that threatens our very future as a Nation and as people. What was unimaginable a generation ago has begun to occur--others are matching and surpassing our educational attainments" (National Commission on Excellence, 1983).

This influential national report may be the first instance where the IEA studies had a direct and lasting effect on U.S. educational policy. Its impact was to gain the attention of politicians from high level administration officials to members of Congress. The reports of comparisons that emanated from the Second International Study in Mathematics (SIMS) later, in 1986, were widely cited in news stories and added fuel to these claims thus increasing interest in international comparisons among U.S. government policy makers. The interest in the international studies led eventually to greater financial support for U.S. participation in comparative studies and also to financial support for the data collection and management activities conducted outside of the United States that are essential for the international comparisons, such as developing questionnaires, centralizing data processing and analysis.

Securing the release of international comparative data from countries for the second round of studies was a complicated process. Although the NCES contract with the University of Illinois called for the production of an international data base in 1981, the IEA did not have a clear method for securing the release of the study results from each participating country. It appeared in 1985 that a coordinated release by the IEA of the survey results for all countries at once was unlikely to happen. The IEA representatives attending the 1986 IEA General Assembly meeting in New Zealand were not strongly in favor of the release of the international comparisons. They were more inclined to use the international framework as a basis for within country analysis and not pursue the more complex problem of between country comparisons (Wolf 1983). Postlethwaite noted the lack of progress in producing rapid results in his opening remarks to the general assembly participants and warned that further delays might jeopardize the funding for the international activities. The difficulties in communication between countries in the 1980's, before the wide use of the electronic internet, greatly increased time delays and confusion by the statisticians about the quality of individual country surveys. However,

with the encouragement of the administrator of NCES, Emerson Elliott, I created a summary table of the average levels of achievement in mathematics for each participating country and gave it to a reporter of the Los Angeles Times who had heard about the study and had called for information. The news story of international achievement differences between countries was repeated in many countries.

The IEA continued to pursue new studies after the success of the SIMS but by 1986, secure funding was not yet assured. The Second International Science Study (SISS) was begun with Malcolm Rossier of Melbourne Australia as the lead. In that year, NCES had no funding available to carry out the appropriate survey field work. But NCES technical staff provided a sample of schools to the study director Willard Jacobson at Teachers College, Columbia University for the American portion of that study. Some support was also provided by the National Science Foundation's office of Assessment (Richard Berry program director). However, the U.S. participation in SISS achieved a school level response rate of less than 50 percent and had to be repeated a year later (by Teachers college, Columbia University) yet the results was an embarrassment to all concerned because of the response rates and irresponsible interpretation of the differences in the two studies. The U.S. participants published the results of both efforts, found a difference, and claimed that the difference was in fact a change over time when in fact the difference was more likely due to the bias in the sample caused by the low response rate. Nevertheless, the press was interested in international studies by this time and SISS received significant national publicity and also criticism for its sampling procedures. This interaction of strong public interest in the studies and the need for funding to support them at a higher level led to discussions within the US government for how to create a stronger support network.

The next study being formed by the IEA by Neville Postlethwaite was to repeat, and improve, the first studies in an international study of

Reading Literacy. The NCES encouraged this study and hosted an international study directors meeting in Washington DC in 1989 to organize that study. Hans Wagemaker and Warwick Elley took leadership roles in that study based on Elley's extensive experience in reading achievement studies in New Zealand. This study was well conducted and provided useful international comparisons and insights into student learning of reading.

Also during this period, some efforts were made to create an international study of writing led by one of the founders of the IEA in the United States, Allen Purves, and U.S. educator Eva Baker. Data were collected in several countries and attempts were made to code and compare country to country in student ability to compose paragraphs. For example, the international committees struggled with standards that would allow for equal standards for lengthy paragraphs of the French with more limited English paragraphs by the Americans. However, after extensive analysis of these efforts by nationally known psychometricians, such as Robert Linn, a decision was made by NCES leadership that cultural differences in composing written language were too great to allow for valid international comparisons.

Although the results of the IEA comparative studies conducted during the 1980's were used in policy arguments and in journals, many policy makers did not trust the data because they were not collected by official governmental agencies to insure standards of statistical procedures. Efforts to publish the survey results in NCES publications met strong resistance from those who believed that the low school and student participation rates meant that the indicators were not representative of the populations they claimed to represent. The IEA studies tended to be conducted by research departments in universities that did not have strong track records of statistical data collection and publication. The response rates to the original sample were below the standards of federal agencies of at least 85 percent. The purposes of the statistical agency were believed to be compromised by publishing the

results of SIMS. A board of nationally known statisticians was created (The Board on International Comparative Studies in Education) at the National Academy of science and was jointly supported by NCES and NSF from 1988 to 2003. The voluntary members of this committee provided guidance to the conduct and interpretation of the international studies. This board was staffed by Dorothy Guilford who had been the Director of NCES in the 1970's. Dr. Guilford was able to engage a strong leadership in statistics and education policy. The board provided guidance to the US data collection and their advice also was sought by other countries. Some of the members of that board included: Norman Bradburn (Chicago), Steven Heineman (World Bank), Andrew Porter (Wisconsin), Marshall Smith (Stanford), Mike Kirst (Stanford), James Guthrie (Berkeley), and Gordon Ambach (Council of Chief State School Officers). The board published a number of useful analyses and procedural guides for conducting comparative studies. That board recommended that the survey results from many of the second round of IEA studies should be submitted for peer review in academic journals but not as official statistics by a federal agency. Consequently, the leaders of the NCES encouraged the creation of a different approach that would be managed by one of the leading assessment companies in the United States.

### **The International Assessment of Education Progress**

While the IEA appeared to be struggling with obtaining reliable international statistics that government policy makers could feel assured were worthy of use in policy discussion, the Education Testing Service (ETS) proposed to conduct a carefully managed brief study of student achievement only strictly for the purpose of establishing the country to country differences. Two surveys were conducted, one in 1988 and another in 1992. The first study included only 5 countries and the second included twenty countries. The surveys collected data on achievement in mathematics and science and on students' attitudes for 9- and

13-year-old students. Unlike the IEA studies, the International Assessment of Education Progress (IAEP) did not attempt to study all aspects of the educational system (such as characteristics of teachers or curriculum) that might account for the country to country differences in student performances. Rather, the purpose of the study was simply to establish whether or not international comparisons could be created with high response levels at all. The study was conducted with funding from both NCES and NSF and resulted quickly in a publication of country rankings that proved to policy makers that international comparisons could be well conducted. The Assistant Secretary of Education for the U.S. Department of Education, Chester Finn, wrote a letter to the education departments of all countries encouraging them to participate in the study. This encouragement led to a greater involvement of other countries in the comparative study. While the ETS study provided quick results and achieved high response rates with a random sample of schools, and thus became a competitor for the IEA, education leaders, such as Leigh Burstein, argued that the IAEP was not as useful in identifying policy levers that could propel further discussion of the education system. Thus, NCES again turned to the IEA for more thoughtful data collection and analysis.

### **Third International Studies**

The full story of the development and production of the Third International Mathematics and Science Study is too complex to report completely for this paper. TIMSS was probably the largest study of education ever conducted in both number of people involved in its design and in the data collection itself. The study began with an idea from David Robitaille (British Columbia) in 1990 that a third series of international comparisons should be created to extend the measurement of the second math study. However, the recent experiences of the IAEP and new efforts of governmental agencies to integrate their own national testing with the international studies led to a different approach that



combined subjects into a single study. This simple idea, first proposed by Gary Philips director of the assessment branch of NCES, created a large complex study because frameworks had not been created to undertake such a study. With funding from both the National Science Foundation and the NCES, international centers for TIMSS were established in Canada at the University of British Columbia, in Germany at the University of Hamburg, in Australia at the Australian Center for Education Research. The center in Canada was moved to Boston College under the direction of Al Beaton and Ina Mullis (both previously had been at ETS) and Michael Martin of Ireland.

Several important innovations for cross-national studies were introduced during the TIMSS study. By 1994, communication between countries through electronic mail was possible. This technological trick allowed for rapid exchange of test instruments and discussion between countries during the development of the test items. Second, Andreas Schleicher a graduate student of Neville Postlethwaite wrote a program that simplified the standardization of country entry of their data into a common data base. This little program improved the standardization and speed of the international studies by a factor of 10. Third, the study attempted to directly address through its framework of design research questions about the causal relationship between what is taught in school and whether it is learned. Fourth, the study included populations at all levels of education: early elementary, late elementary and late secondary. Fifth, the rapid publication of results of the study attracted wide national and international attention to its results. These innovations were the product of many individual intellectual contributions as well as a willing partnership between funding agencies, the international organization, and the 50 participating countries.

### **Individuals that Affected IEA Studies**

I have had the great privilege and honor of meeting and working with hundreds of participants in the IEA studies of Mathematics, Science,

Reading and Video studies over the past 35 years. I will confine my remarks here to those leaders with whom I conducted the business of funding people and projects of the IEA and not with those who actually carried out the important local studies or the international data management although these performers are important for the outcome of the studies. The most involved IEA leaders include Torsten Husén (Sweden), Neville Postlethwaite (Germany), John Keeves (Australia), Gilbert DeLandshere (Belgium), Tjeerd Plomp (Netherlands), Hans Wagemaker (New Zealand), Claire Burstall (England), Seamus Hagerty (England), and David Robitaille (Canada). Some of the American leaders in IEA studies include Richard Wolf (Columbia), Alan Purves (U. Illinois), Steven Heyneman (World Bank), Kenneth Travers (U. Illinois), Willard Jacobson (Columbia) and Rodney Doran (Buffalo), Leigh Burstein (UCLA), William Schmidt (Michigan State), and Albert Beaton (Education Testing Service and Boston college). Dr. John Anderson of the University of Chicago was the first person to receive federal funding from the Office of Education to conduct an international study of mathematics in 1964. He was a professor of comparative education and he managed to receive the award that continued the earlier work of the IEA after it left the UN organization that had founded it.

These leaders helped to pursue the idea that comparisons of student performance could be made at all. Some of the critics of the first set of studies such as Hans Freudenthal (Netherlands), a mathematician, would have stopped government support of the studies if the responses of Postlethwaite were not so well presented and received (Freudenthal 1978). Postlethwaite was a high energy level manager and administrator who liked to exchange ideas with strong minded intellects. Husén was indeed an intellect with a great dedication to understanding the elements of educational practices and presenting them to world leaders in a convincing manner. Husén was provided a number of opportunities to simplify the results of the international studies in order to please administrators, but he refused to do that.

While his attention to details and the complexities of educational organizations remained intellectually sound, he was not effective in raising funds for his projects with this approach. Postlethwaite took a more pragmatic view toward funding agencies and used his phone frequently to discuss staff issues, organizational arrangements, statistical analyses plans, and even arguments about weighting and adjustments for non-response all of which concerned the funding agencies. Postlethwaite was effective in gaining support for studies that would bring together a large number of countries with the least amount of resources. His model of organization grew from the relationships of research centers in universities around the world. He and his supporters (such as Leigh Burstein of UCLA) believed that the best way to achieve cooperation of schools and to study countries was by keeping a low profile at the international organizations. They believed that once the "league tables" was published that the schools would be less likely to participate in studies.

Tjeerd Plomp was an effective leader in a period of transition. He was inclusive in his organizational arrangements and committee assignments. He managed a growing organization during the TIMSS early years and learned along with others about how to present a complex study to the world at large in an effective manner. He recognized the possibility of organizing the public release of TIMSS in such a way as to maximize, rather than minimize, the public attention that it received.

The creation of the internet greatly improved the quality of international studies. The Communication between countries was slow and uncertain before the 1990's and the IEA did not have effective management of the sampling and data collection procedures. However after electronic communication through the internet became a standard means for communication the easy exchange of large data files improved the ability of project managers to conduct large studies across large distances.

The American leaders who participated in the Board on International

Comparative Studies in Education (BICSE) all had an important influence on the international leadership through their advice and involvement in funds management, analysis plans, staff selection, and interpretation of statistical comparisons of complex systems. Emerson Elliott of NCES should be recognized for his continuing support of the IEA studies but with care and concern for the quality of the studies.

Some recognition should be given to three loud critics of the publication of international comparative studies because their criticisms led to the need for improvements in the data collection and reporting procedures: Iris Rotberg, Gerald Bracey, and David Berliner. Iris Rotberg, who worked then at the National Science Foundation, believed that the international comparisons presented an overly negative view of the quality of the public school system. She criticized the comparisons for not being fair because the student populations were not equally selected. While she may not have had the technical details correct, her public presentations tended to bring greater attention to how comparisons between countries could be interpreted and the responses by other researchers to these concerns helped to clarify their meaning to a skeptical audience of listeners. Gerald Bracey who published his annual "Condition of Public Education" in the Phi Delta Kappa magazine tended to find reasons for not believing that international comparisons presented a fair indictment of American public schools. These views were also presented in books by David Berliner and Bruce Biddle (Berliner, & Biddle, 1995). The manufactured crisis. New York: Addison- Wesley) that argued that the interpretation of US achievement compared with other countries overstated the problems with the education system itself and brought unnecessary reform attempts to the public school system. These critics contributed to the growing improvement of the funding for the studies by bringing to light their value in public discourse. If the critics believed that the publication of international results were giving attention to the wrong aspects of education, they only succeeded in giving greater emphasis to the background and quality of the studies.

## General Philosophy of IEA

My own role has been as a consistent supporter, intellectual advisor, financial provider, and sometimes critic of the IEA. The IEA I first met in 1981 was a research organization struggling with creating conceptual frameworks for making comparisons between whole societies. That problem had not been solved. The leaders, as can be seen in the publications from the first set of studies and the developers of the Second series of Math and Science studies, avoided making comparisons between countries because they were concerned that the comparisons were invidious. The publications of the first reports, for example, did not include league tables. The originators of the IEA believed that the main effort of the international collaboration was to improve within country analysis of its own education system by taking an objective approach to the measurement of student performance (see Wolf 1983). For example, Torsten Husén called this the “comparison of the incomparable” (Husen 1987). The philosophy of the group was to allow each country to represent its findings in a way compatible with its own assumptions and present results of individual items only. Scales of total achievement were not created centrally until TIMSS developed scales directed by Al Beaton who had created the methodology for multiple estimators for the U.S. NAEP data collection. Beaton also introduced a large number of comparisons tables, after long discussion and debate arguing that the presentation of “league” tables was not inherently bad, but that it was best to present many separate ranked tables on different aspects of achievement. Therefore, the subsequent TIMSS reports have contained large tables of country comparisons.

The methods of communicating between countries provided major challenges to IEA for getting the work of statistical data together for reliable measurement. The results of the first IEA studies during the 1960 and 1980's took several years to analyze and publish because of the technology available at that time. Moreover, the accessibility to

data files by other researchers was very limited in those years. In the first IEA studies, and in the initial stages of new studies in mathematics and science, some of the best educational policy makers and educational researchers in the world were recruited to participate in establishing frameworks and conducting studies. The second and third round of IEA studies greatly improved the access to data through internet communication and the large studies continued to attract researchers who maintained high quality of data collection and reporting.

Over the past 50 years the IEA has changed from a small intellectual body conducting research to a large organization prepared to meet the demands of countries that seek objective measures of their student performance for accountability purposes. The demands of data collection and fund raising to continue existing large studies consume significant amounts of energy for the association members. The intellectual excitement of searching for new insights into educational practices that characterized the originators of the IEA may have been replaced by the need for continuing the infrastructure of a large international organization. On the other hand, the quality of the published data matches that of the best statistical organizations in governments and international organizations around the world.

The IEA studies have introduced many important empirical indicators of previously abstract educational practices regarding the curriculum, teaching practices, and student learning. The IEA studies have provided policy makers with well designed conceptual frameworks for student assessment and educational practices believed to be related to effectiveness of schools as learning centers. For example, the opportunity to learn measures were an international invention intended to explain country differences but that became a within country concern. The IEA proved that international comparisons of the incomparable (said Husén) could be accomplished. These studies have changed the discussion of policy from participation rates in school to understanding learning itself through analysis of curriculum and teaching. The

impact of the IEA on policies of inclusion of minority populations in the state educational system is less notable. The first publications of Torsten Husén were about the development of talent in countries where individuals with ability were being overlooked by an elitist school system. These topics have not been addressed as directly in the later analyses.

The IEA contributed to the development of talent in educational research through broadening the participation of empirical data collection on student achievement to many more countries. Many more academic publications of educational practices have been produced as a result. The IEA publications introduced the notion of 'opportunity to learn' as a conceptual framework and as a way of thinking about the role of educational policy. The studies directed research toward the processes of curriculum and teaching rather than administrative practices of schools as had been the emphasis before international studies. The publicity of results on student achievement levels in mathematics, science, reading, and computer technology has raised the level of discussion about student achievement in the minds of the general public and the politicians of our age. In fact, the growth of the industry of educational testing could be the result of the success of the international comparisons in giving public attention to the status of each country's school system. Administrators, politicians, and researchers alike have learned to turn to statistical analysis for information about the school system before embarking on major reforms.

### **Suggestions for Improvement**

The collection of the same statistics time and time again may be useful for national descriptions, but they may not address new significant issues that countries face as their systems progress over time. The IEA should use these data and conversations with policy makers to define some enduring educational issues that would drive future studies. For example, the trends are showing that some countries are declining in

achievement levels. What are the underlying factors for decline?

Getting access to complex data systems is possible today because of access through the internet. The OECD has created a much more flexible and creative procedure for accessing their survey data than has the IEA. The IEA needs to continue to prepare their data sets to make them easily available to researchers and policy makers to insure that their survey data are used. Hard paper publications are no longer sufficient.

The IEA once sponsored fellows to live in another country and participate in research on a study. The IEA should expand its training opportunities by seeking funding for many more such international exchanges. Bring young researchers into the analysis of data through methods such as fellowships and small grants.

### **Memorable Moments**

Of course, travel to new cultures always leads to personal insights and experiences. The IEA is for some people a travel excuse. For me however, more important has been the intellectual discussions and insights gained from those in other countries. However, some of the early travels will remain with me as memorable. For example, at my first IEA General Assembly meeting in New Zealand in 1986 the Maori tribe provided a reception in Auckland that included an evening of speeches and dances that clearly were not from the Hollywood culture so common in the entertainment world. This meeting was all the more dramatic because I was attempting to convince members of the IEA that it would be in their own interest to release the results of the assessment results in a simple format of comparisons of average percentage correct (this was before IRT models were easily computed).

Another memorable event was the meeting of the TIMSS group in Italy to discuss the final release of the TIMSS results. The meeting was held in a villa Falconieri in Frascati far removed from the busy city of



Rome and in an area where the details of analysis results could be conducted quietly and privately. The villa provided an other-worldly experience since it was separated from the culture of modern Rome and seemed to be more relevant for the Rome of a thousand years ago.

I have met the most interesting educational researchers in the world through the IEA meetings in The Netherlands, Germany, Italy, Switzerland, Spain, England, Norway, Sweden, Thailand, France, Canada, USA, Czechoslovakia, Chile, and New Zealand. And I have made friends in more than 50 countries who share my interest and enthusiasm for international comparative research. Many of these administrators and researchers are now active in other international studies such as PISA by the OECD demonstrating that the IEA is the father of comparative statistics.

In my role as funder and researcher I have supported the use of Federal funds to administer the IEA studies in the United States and for the international organization itself. I used IEA data in discussions of policy and direction with in the National Science Foundation. The NSF uses the IEA data in the elementary and secondary education chapter of the biannual report "Science and Engineering Indicators" that is published by the prestigious National Science Board in every edition since 1996 including the recent 2010 publication. <http://www.nsf.gov/statistics/seind10/>. This report is widely read by scientists and administrators and policy makers in the government.

## Conclusion

The IEA will continue for many years to share with other international organizations the responsibility of objective measurement of educational outcomes. It will provide training in needed places and will invent new approaches to data exchange. The information on trends that are currently being collected will someday become recognized as a major contribution to the understanding of the large educational systems that countries manage.

## References

- Berliner, D. C. & Biddle, B. J. (1995). *The manufactured crisis (myths, fraud, and the attack on America's public schools)*. Reading, MA: Addison-Wesley Publishing.
- Burstein, L. (1989). *Second International Study of Mathematics* Vol. 3. London: Pergamon Press.
- Freudenthal, H. (1978). *Weeding and sowing: Preface to a science of mathematical education*. Amsterdam: Reidel.
- Lerner, B. (1982). American Education: How are we Doing? *The Public Interest*, 69, 59-82.
- National Commission on Excellence in Education. (1983). *A nation at risk: The imperative for educational reform*. Washington, DC: U.S. Government Printing Office.
- Wolf, R. (1977). *Achievement in America*. New York: Teachers College Press.
- Wolf, R. (1983). Are Standards in U.S. Schools Really Lagging Behind Those in Other Countries? *Phi Delta Kappan*, 64, 455-461.





APPENDIX  
**AUTHOR BIOGRAPHIES**



**Ainley, John** (BSc in chemistry and PhD in education, both from the University of Melbourne) is a Principal Research Fellow and former Deputy CEO at the Australian Council for Educational Research. He represented Australia in the General Assembly of the IEA from 2003 to 2010. He was the Australian national research coordinator for the Second Information Technology in Education Study (SITES) Module 2 (which investigated innovative pedagogical approaches using ICT) and SITES 2006 (which surveyed schools and mathematics and science teachers on ICT use and ICT-related approaches to teaching). He was international coordinator of the International Civic and Citizenship education Study (ICCS) from 2006 to 2011 and is currently coordinator of the International Computer Information Literacy Study (ICILS).

*Address:* ACER, 19 Prospect Hill Rd, Private Bag 55, Camberwell, Victoria 3124, Australia

*Email:* ainley@acer.edu.au

Co-author of chapter 8, "The International Civic and Citizenship Education Study (ICCS)", and chapter 14, "The Contribution of IEA Research Studies to Australian Education".

**Ambach, Gordon** represented the United States in the IEA General Assembly and served on the Standing Committee from 1990 to 2001. Career in education administration: U.S. Office of Education, 1961-64; New York State Education Department, 1967-87, including Commissioner of Education, 1977-87; Executive Director, Council of Chief State School Officers, 1987-2001. Board memberships: Smithsonian Institution National Board, National Academy of Sciences Education Board, Wallace Foundation, Center for Naval Analysis, and New Haven Symphony. Retired.

*Address:* Residence 274 Ogden St. New Haven, CT, 06511, USA.

*Email:* ambachg@aol.com.

Author of chapter 20, "The IEA of 1990-2001: a 50th Year Reflection on Transformation and US Participation".

**Blahova, Viera** is a state adviser for ICT in Education in the Department of



Informatics in the Ministry of Education, Science, Research and Sports of the Slovak Republic. After her studies at the Comenius University in Bratislava, she worked as teacher, educator and researcher. In 2004 she joined the Ministry of Education. She was a national coordinator for SITES (Second Information Technology in Education Study) in Slovakia. She was a coordinator of the international project IDWBL (Innovative didactics for web-based learning). She is an expert on policies and practices in the field of ICT in Education in Slovakia. She is a member of the Steering Committee for European Schoolnet and is on the Steering Committee for the e-skills week campaign of the European Commission.

*Address:* Ministry of Education, Science, Research and Sports of the Slovak Republic, Stromova 1, 813 30 Bratislava, Slovak Republic.

*Email:* viera.blahova@minedu.sk

Co-author of chapter 19, "IEA Experiences from Latvia, Lithuania, Slovak Republic and Slovenia".

**Brassói, Sándor:** MA in Biology and Geography at University Eötvös



Loránd, Budapest. Between 1986 and 1998, teacher, mentor teacher and teacher trainer. Deputy school principal in Városmajori Gimnázium, an academic secondary school in Budapest. From 1998 to 2010, deputy head of Department for the Development of Public Education in the Ministry of Education. From 2010 head of Department for Public Education in the State Secretariat for Education of the Ministry of Human Resources.

In his government functions he has been responsible for research and development in public education, including legislation, planning and coordination of national and international assessment and evaluation programmes in public education. He has been representing Hungary in the IEA General Assembly since 2003.

*Address:* Ministry of National Resources, State Secretariat for Education, 10-14 Szalay utca, Budapest, H-1051 Hungary

*Email:* sandor.brassoi@nefmi.gov.hu

Co-author of chapter 18, "How IEA Influenced the Education System in Hungary".

**Brusselmans-Dehairs, Christiane** worked since 1970 as a researcher at the University of Gent on many IEA studies. She was a researcher in the Six Subject Study (1970–71), Second International Mathematics Study (1980–82) and Computer in Education Study (1989, 1992); and National Research Coordinator for the IEA Trends in Mathematics and Science Project (1995, 1999, and 2003). She served as IEA General Assembly representative for the Flemish Community (1992–1997) and was elected in September 1995 as member of the IEA Standing Committee, representing both the Flemish and Walloon communities of Belgium. She has authored numerous IEA-related publications and reports. Christiane retired in 2007.



*Address:* Lauwstraat 81, B9051 Sint-Denijs-Westrem, België

*Email:* [christiane.dehairs@telenet.be](mailto:christiane.dehairs@telenet.be)

---

Roundtable participant in chapter 21, “Roundtable at the 49th General Assembly, Berlin 2008”.

**Cariola, Leonor** has worked in the Ministry of Education since 1998, leading the International Studies Team. She is a sociologist (Pontifical Catholic University of Chile) and holds a PhD (University of Wales). For over 25 years she has worked as a researcher in topics related to education and has been a consultant in different countries in regard to secondary education, curriculum and assessment. At the Ministry, within the student national evaluations at the National Assessment Unit of the Ministry of Education of Chile, she was the national research coordinator for CIVED and for TIMSS 2003. She was also member of the Project Advisory Committee for the ICCS study. Currently, she is working part-time as a consultant for research and development in the Ministry.



*Address:* SIMCE Ministry of Education. Alameda 1146, Sector B, Piso 8, Santiago de Chile.

*Email:* [leonor.cariola@mineduc.cl](mailto:leonor.cariola@mineduc.cl)

---

Co-author of chapter 15, “Chilean Participation in IEA Studies”.



**Čepič, Mitja V.** is a researcher in the Department of Applied Epistemology at the Educational Research Institute in Ljubljana. His main area of work is the analysis of discourse pertaining to knowledge and/or competence in curricula, textbooks and other educational texts. He has also collaborated on studies of civic education in Slovenia.



*Address:* Pedagoški inštitut / Educational Research Institute, Gerbičeva 62, SI-1000 Ljubljana, Slovenija;

*Email:* mitja.cepic-vogrincic@pei.si

Co-author of chapter 19, "IEA Experiences from Latvia, Lithuania, Slovak Republic and Slovenia".

**Covacevich, Catalina** has an MSc in Social Psychology, London School of Economics and Political Science; Licenciante in Psychology, Pontifical Catholic University of Chile. Catalina worked for the Catholic University of Chile developing instruments for the Chilean national assessments for teachers. She then worked for the Chilean Ministry of Education, where she was the national research coordinator for the IEA's ICCS study from 2007 to 2010. In the same period she was also the national coordinator for the Latin-American initiative SREDECC (regional system of assessment and development of civic competencies). She is currently back at the Catholic University, coordinating the design and development of the tests for UNESCO'S Latin-American study TERCE (third regional explanatory and comparative study).



*Address:* Centro de Medición MIDE UC, Escuela de Psicología, Pontificia Universidad Católica de Chile Vicuña Mackenna 4860 Macul, Santiago Chile.

*Email:* ccovacep@uc.cl

Co-author of chapter 15, "Chilean Participation in IEA Studies".

**Dukynaitė, Rita** has an MSc in Philology (Vilnius University) and a PhD in



Social Sciences. She is an Associate Professor of Education and the Deputy Head of Strategic Planning and Analysis Division of the Ministry of Education and Science of the Republic of Lithuania. Rita Dukynaitė has represented Lithuania in the General Assembly of the International Association for the Evaluation of Educational Achievement (IEA) since 2004, and has been a member of IEA's Standing Committee since 2011.

*Address:* Andriaus Volano 2/7, LT-01516 Vilnius, Lithuania

*http://*[www.smm.lt](http://www.smm.lt);

*Email:* Rita.Dukynaite@smm.lt

---

Co-author of chapter 19, "IEA Experiences from Latvia, Lithuania, Slovak Republic and Slovenia".

**Fraillon, Julian**, BA, Dip. Ed. Dip. Mathematics Ed., Postgrad. Dip.



Mathematics and Mathematics Ed (University of Melbourne). Julian is the Research Director of the ACER Assessment and Reporting Research Program: Mathematics and Science. He is the research director of the International Computer and Information Literacy Study (ICILS) commissioned by the International Association for the Evaluation of Educational Achievement (IEA); assessment manager of the Australian National Assessment Program (NAP) ICT Literacy 2011; project director of the NAP Civics and Citizenship Assessment 2010; and manager of cognitive test development and the coordinator of the Asian Regional Module for the IEA International Civics and Citizenship Education Study 2009 (ICCS). Recently, Julian has also managed the assessment content of both the NAP ICT Literacy 2008 and 2005 and Civics and Citizenship, 2007 and 2004 projects.

Julian has worked both on the substance and as a consultant reviewer of assessment programs for both academic and non-academic outcomes of schooling across Australia and internationally.

*Address:* ACER, 19 Prospect Hill Rd, Private Bag 55, Camberwell, Victoria 3124, Australia

*Email:* fraillon@acer.edu.au

---

Co-author of chapter 8, "The International Civic and Citizenship Education Study (ICCS)".

**Gubler, Johanna** (Psychologist, Pontifical Catholic University of Chile, graduated in 1997). Johanna worked on the development of student assessment instruments and achievement standards for the student national evaluations at the National Assessment Unit of the Ministry of Education of Chile (2004 to 2008). In 2009 she joined the International Studies team at the Ministry of Education and became the National Research Coordinator (NRC) for the Trends in International Mathematics and Science Study TIMSS 2011. Currently she is the Coordinator of the International Studies team at the Ministry of Education



*Address:* SIMCE Ministry of Education. Alameda 1146, Sector B, Piso 7, Santiago de Chile.

*Email:* Johanna.gubler@mineduc.cl

Co-author of chapter 15, "Chilean Participation in IEA Studies".

**Howie, Sarah**, BA (Stellenbosch), BA Hons (Cape Town), M.Ed (Witwatersrand), PhD (Twente)) is the Director of the Centre for Evaluation and Assessment, Professor in the Faculty of Education at the University of Pretoria. She has extensive experience in IEA studies and has served as National Research Coordinator for studies such as TIMSS 1995, TIMSS 1999, PIRLS 2006 and SITES 2006, and was involved in SITES M1 and SITES M2; she serves on the international Questionnaire Development Group for PIRLS 2011, was a member of the Scientific Committee for the IEA's International Research Conference 2010 and has also contributed to a number of publications in this research domain. She serves on inter alia the Assessment and Statistical committee for the South African Quality Assurance and Certification council, Umalusi, standardising the national examination results annually, and served on the South African Ministerial Committee on Learner Retention. Her doctoral work on the TIMSS 1999 won the National Science and Technology Forum's prestigious prize for Most Innovative Research 2003.



*Address:* Centre for Evaluation and Assessment, AIS building - Room 2.9, Faculty of Education, University of Pretoria - Groenkloofcampus, Pretoria 0002, South Africa

*Email:* sarah.howie@up.ac.za

Author of chapter 13, "The Involvement of African Countries in the IEA Studies over 50 Years".

**Japelj, Pavešič, Barbara** has been the National Research Coordinator for



TIMSS in Slovenia since 1999. She was part of the national research team on the first TIMSS study in 1995. She is employed at the Educational Research Institute where she is working on different research projects related to mathematics education. Her main research interest is the use of modern statistics methods of data mining and network analysis in educational research. She also writes textbooks for elementary school mathematics and educates teachers in the context of professional development programmes.

*Address:* Pedagoški inštitut / Educational Research Institute, Gerbičeva 62, SI-1000 Ljubljana, Slovenija

*Email:* barbara.japelj@pei.si

---

Co-Author of chapter 19, "IEA Experiences from Latvia, Lithuania, Slovak Republic and Slovenia".

**Justin, Janez**, PhD is head of the Department of Applied Epistemology at



the Educational Research Institute in Ljubljana. His main research interests are didactic discourse and communication, textbook research and research on civic education. He is also Slovenia's representative in IEA's General Assembly and a member of the Slovenian Council for Evaluation of Education at the Ministry of Education and Sport.

*Address:* Educational Research Institute (ERI), Gerbiceva 62, SI-1000 Ljubljana, Slovenia.

*Email:* janez.justin@guest.arnes.si

---

Co-author of chapter 19, "IEA Experiences from Latvia, Lithuania, Slovak Republic and Slovenia".

**Kádár-Fülöp, Judit**, MA in English and Russian linguistics and literature



University Eötvös Loránd, Budapest, CSc in educational sciences Hungarian Academy of Sciences, Dr.habil. University of Debrecen. From 1970 to 1991 researcher at the National Institute for Education. From 1991 to 2000 deputy head of department in the Ministry of Education, responsible for development projects in public education and educational statistics. 2001-2003 analyst, OECD Secretariat, Paris. From 2004 senior policy advisor to the Ministry of Education.

During her time with IEA, she served as national research coordinator in the 1970 Six Subject Survey for English as a Foreign Language, in the 1981-1984 Written Composition Study and in the 1991 Reading Literacy Study. She was a member of the International Steering Committee for the Written Composition Study.

*Address:* Ministry of National Resources, State Secretariat for Education, 10-14 Szalay utca, Budapest, H-1051 Hungary

*Email:* Judit.Kadar@nefmi.gov.hu

Co-author of chapter 18, "How IEA Influenced the Education System in Hungary".

**Kangro, Andris**, Dean of the Faculty of Education, Psychology and Art at



the University of Latvia since 1998. He has a Doctorate in Physics, and is a professor in Education Management. He has been the IEA General Assembly member for Latvia since 1993, and was an IEA Standing Committee member (1999-2002). He is head of the IEA national research centre at the University of Latvia (1993-1999) and has been the director of the Institute for Educational Research of the Faculty of Education, Psychology and Art at the University of Latvia since 1999. He has been the OECD PISA National Project Manager since 1998 and he was National Research Coordinator for the IEA CIVIC study and IEA ICCS. He has published widely.

*Address:* Faculty of Education, Psychology and Art of University of Latvia, Jūrmalas gatve 74/76, Rīga, LV-1083, Latvia.

*Email:* Andris.Kangro@lu.lv

Co-author of chapter 19, "IEA Experiences from Latvia, Lithuania, Slovak Republic and Slovenia".

**Keeves, John P.** (BSc Hons in physics, MEd in education, PhD in sociology)



is Adjunct Professor at the University of Adelaide, Australia, and Honorary Life Member of the IEA. He represented Australia at the IEA General Assembly as Associate Director and Director of the Australian Council for Educational Research from 1972-84. He was Visiting Scholar from 1971-72, and Visiting Professor from 1987 to 1989, at the Institute of International Education, University of Stockholm and Chair of the Flinders University Institute of International Education from 1998 to 2006. He was also Chair of the International Steering Committee of the Second IEA Science Study (SISS) from 1978-1989.

*Address:* P.O. Box 1883, Burnside, S.A.5066, Australia

*Email:* john.keeves@flinders.edu.au

Author of chapter 1, "IEA - From the Beginning in 1958 to 1990", and co-author of chapter 10 "The Relationship of IEA to some Developments in Educational Research Methodology and Measurement during the Years from 1962 to 1992", and chapter 14, "The Contribution of IEA Research Studies to Australian Education".

**Lagos, Ema**, Sociologist at University of Chile. Ema worked for 10 years



developing political and social research projects in non-government research institutes and private companies (CERC and MORI). In 1998 she started to work for the national evaluation system in the Ministry of Education, first in the Civic Education Study and later, taking part in different international educational studies, as member of an international studies team. She also was assistant to the Technical National Coordinator in the national evaluation system (2005-2006), and she has been the national coordinator of the PISA project since 2007.

*Address:* SIMCE Ministry of Education. Alameda 1146, Sector B, Piso 7, Santiago de Chile.

*Email:* ema.lagos@mineduc.cl

Co-author of chapter 15, "Chilean Participation in IEA Studies".

**Lehmann, Rainer**, doctorate in theology and PhD in education, both



University of Hamburg), professor emeritus of education (educational research methods) at Humboldt University, Berlin, Germany. He represented Germany in the General Assembly of the International Association for the Evaluation of Educational Achievement (IEA) from 1990 – 2007.

During his time with IEA, he worked as National Coordinator in several IEA projects (Written Composition, Reading Literacy, TIMSS) and served as International Coordinator of the Civic Education Study (CivEd).

*Address:* Humboldt University, Philosophische Fakultät IV [Sitz: Georgenstr. 38], Unter den Linden 6, DE-10099 Berlin, Germany

*Email:* rlehmann@educat.hu-berlin.de

Author of chapter 17, "The Impact of IEA on Educational Policy-Making in Germany" and chapter 22, "The Scientific Contributions of Torsten Husén and Neville Postlethwaite to the Development of International Comparative Research on Educational Achievement".

**Leimu, Kimmo A.I.**, Teacher's training at Helsinki Teachers College, 1959.



Ed-Lic. (Psychology), University of Jyväskylä, 1972. Teaching experience at primary, secondary, tertiary and university levels. Joined IEA in 1967. National Research Coordinator for several IEA projects. Finnish representative at the IEA General Assembly from 1970s to 1996. Member of the IEA Standing Committee, 1983-87; Training Committee 1983-85; Policy and Procedures Committee 1983-85; Publications and Editorial Committee 1986-96; and editor of the IEA newsletter 1985-1988. Leader of the research group "Comparative Evaluation

of Education" at the Institute for Educational Research, University of Jyväskylä, from 1994–1999. Retired in 1999. Honorary member of IEA, 2009. IEA-type research applications as UNESCO Adviser at the Ministry of Education, Lesotho, 1976-78, and in other educational R&D projects, including Botswana, Lesotho, Zambia, Palestine and Cambodia, between 1976 and 2001. Finnish representative in the OECD/CERI International Education Indicators Project (INES) Network A: Student learning outcomes (1990-1999). Since retirement has functioned as School Quality Monitor for projects OECD/INES/PISA: (Programme for International Student Assessment), in 2000, 2003, and 2009, as well as in 2009 for IEA/ICCS – International Civic and Citizenship Education Study, and in 2011 for IEA/TIMSS 2011 – Trends in International Mathematics and Science Study.

*Address:* Ollintie 4 B 10, 40270 Palokka, Finland.

*Email:* kimmo.leimu@luukku.com

Author of chapter 26, "50 Years of IEA: A Personal Account"

**Leung, Frederick Koon-Shing**, is a professor in mathematics education



at the Faculty of Education of the University of Hong Kong. Professor Leung obtained his BSc, Cert.Ed. and M.Ed. from the University of Hong Kong, and subsequently his PhD from the University of London Institute of Education. His major research interests are in the comparison of mathematics education in different countries, and in the influence of different cultures on teaching and learning. He is principal investigator of a number of major research projects, including the Hong Kong component of the Trends in International Mathematics and Science Study (TIMSS), the TIMSS Video Study, and the Learner's Perspective Study (LPS). Professor Leung was a member of the Executive Committee of the International Commission on Mathematical Instruction (ICMI) and the Standing Committee of the International Association for the Evaluation of Academic Achievement (IEA). He also has served as a consultant for the World Bank and UNESCO.

*Address:* Faculty of Education, University of Hong Kong, Pokfulam Road, Hong Kong  
*Email:* frederickleung@hku.hk

Author of chapter 16, "The Significance of IEA Studies for Education in East Asia".

**Lietz, Petra**, B.Ed (Hamburg), M.Ed (Flinders), CTEFLA (Cambridge),



M.Acc (CQU), PhD (Flinders), first came into contact with IEA in early 1988 when she started to work for Professor T. Neville Postlethwaite at Hamburg University, where the International Coordinating Centre for the Reading Literacy Study was located and the data processing and analyses for the Second International Science Study were undertaken. For her PhD, which examined changes in reading comprehension across cultures and over a 20-year period using IEA data, she received IEA's Bruce Choppin award. More recently, she assisted with IEA's workshop series and resulting publication Educational Issues in the Middle East North Africa Region. She has held academic appointments in Australia and Germany and is now Principal Research Fellow in the Adelaide office of the Australian Council for Educational Research.

*Address:* ACER, Level 3, 97 Pirie Street, Adelaide, S.A. 5000, Australia  
*Email:* lietz@acer.edu.au

Co-author of chapter 10 "The Relationship of IEA to Some Developments in Educational Research Methodology and Measurement during the Years from 1962 to 1992", and chapter 14, "The Contribution of IEA Research Studies to Australian Education".



**Lockheed, Marlane E.** (PhD in international development education,



Stanford University), World Bank official (retired). From 1972-1985, she directed research on gender, education and testing at the Educational Testing Service in Princeton, NJ. From 1985-2004 at the World Bank, she was a research sociologist, regional education sector manager, interim director of education and head of the World Bank Institute's evaluation group. In the 1990s, she oversaw World Bank grants to IEA, was the World Bank's observer at IEA General Assembly meetings, and served on the U. S. National Academy of Science's Board for International and Comparative Studies in Education. She was a divisional president and vice-president of the American Education Research Association, 1983-85, and Chair of American Educational Research Association's (AERA) Special Interest Group on International Studies, 1991-93.

*Address:* 561 Lake Drive, Princeton, NJ 08540, USA

*Email:* mlockheed@verizon.net

Author of chapter 28, "Reflections on IEA from the Perspective of a World Bank Official".

**Lokan, Jan,** BA in mathematics, GradDipEd (Adelaide), PhD in education



(Ottawa, Canada), FACE, is a former Deputy Director of the Ottawa Board of Education Research Branch and former Assistant Director of the Australian Council for Educational Research. At ACER she managed several large-scale assessment studies within Australia before becoming the National Research Coordinator for TIMSS in 1993. She co-authored three national reports for that study, one for each population, as well as contributing to the international training team for the performance assessment component. She was a member of the international science expert panel until 2002. She established and managed Australia's participation in both the Mathematics Video Study and the Science Video Study, for which she co-authored the national reports. She presented several TIMSS-based papers at international conferences and organised the program for the inaugural ACER Research Conference in 1997, on the topic of Raising Australian Standards in Mathematics and Science: *Insights from TIMSS*.

*Address:* PO Box 515, McLaren Vale, S.A. 5171, Australia

*Email:* lokan@acer.edu.au

Co-author of chapter 14, "The Contribution of IEA Research Studies to Australian Education".

**Loveless, Tom**, is a Senior Fellow in Governance Studies at the Brookings Institution in Washington, D.C. He is also a member of the Task Force on K-12 Education at Stanford University's Hoover Institution. Loveless's research focuses on education policy. His most recent book is *Lessons Learned: What International Assessments Tell Us About Math Achievement*. From 1979 to 1988, Loveless was a classroom teacher in his hometown of Sacramento, California. After receiving his PhD from the University of Chicago in 1992, Loveless joined the faculty of Harvard University's John F. Kennedy School of Government, where he served as both assistant and associate professors of public policy until 1999. Loveless has served on advisory boards of the *American Journal of Education* and *Educational Evaluation and Policy Analysis*. Since 2004, Loveless has represented the United States at the General Assembly of the International Association for the Evaluation of Educational Achievement.

*Address:* 1775 Massachusetts Avenue, NW Washington, DC 20036-2188 United States

*Email:* [tloveless@brookings.edu](mailto:tloveless@brookings.edu)

Roundtable participant in chapter 21, "Roundtable at the 49th General Assembly, Berlin 2008".

**Masters, Geoff**, BSc, Med (UWA), PhD (Chicago), FACE, FACEL, is Chief Executive Officer and a member of the Board of the Australian Council for Educational Research (ACER) – roles he has held since 1998. He has a PhD in educational measurement from the University of Chicago and has published widely in the fields of educational assessment and research. Professor Masters has served on a range of bodies, including terms as founding President of the Asia-Pacific Educational Research Association; President of the Australian College of Educators; Chair of the Technical Advisory Committee for the International Association for the Evaluation of Educational Achievement (IEA); Chair of the Technical Advisory Group for the OECD's Programme for International Student Assessment (PISA); member of the Australian National Commission for UNESCO; and member of the International Baccalaureate Research Committee. Professor Masters was the recipient of the Australian College of Educators' 2009 College Medal in recognition of his contributions to education.

*Address:* ACER, 19 Prospect Hill Rd, Private Bag 55, Camberwell, Victoria 3124, Australia

*Email:* [masters@acer.edu.au](mailto:masters@acer.edu.au)

Author of chapter 14, "The Contribution of IEA Research Studies to Australian Education".

**Montie, Jeanne** (MA and PhD in Developmental Psychology from Case



Western Reserve University in Cleveland, Ohio), Senior Research Associate (retired) at HighScope Educational Research Foundation, Ypsilanti, Michigan, U.S.A. During her time at HighScope she served as coordinator for U.S. data collection and senior writer for the IEA Preprimary Project. She also served as a co-investigator for the 40-year follow-up of the Perry Preschool Study.

*Address:* 3151 Daleview Dr., Ann Arbor, MI 48105, USA.

*Email:* jmontie157734mi@comcast.net

---

Author of chapter 7, "The Growth and Development of the Preprimary Project".

**Ortiz, Marcela** (Sociologist, Pontifical Catholic University of Chile,



graduated 1985) Marcela worked for the National Commission for Scientific and Technological Research at the Ministry of Education of Chile, in its Advanced Human Capital Program leading the area of monitoring and aiding Chilean scholars. In 2009 she moved to the National Assessment Unit of the Ministry of Education and became assistant to the International Studies team. Later, in 2010 she took part as National Research Coordinator (NRC) for the International Civic and Citizenship Education Study ICCS 2009. Currently, she serves as NRC for ICCS 2009 and as NRC for the International Computer and Information Literacy Study ICILS 2013.

*Address:* SIMCE Ministry of Education. Alameda 1146, Sector B, Piso 7, Santiago de Chile.

*Email:* marcela.ortiz@mineduc.cl

---

Co-author of chapter 15, "Chilean Participation in IEA Studies".

**Papanastasiou, Constantinos**, professor at the University of Cyprus since



1993, retired in 2008; he has also taught as a visiting professor at Michigan State University, at the University of Athens and at the University of Crete. He obtained his PhD from Kent State University, United States and is specialized in educational research, statistics, measurement, and evaluation. He has served as the representative for Cyprus to the IEA General Assembly (1992-2008) and member of the IEA Standing Committee (1999-2001). He also served as a national research coordinator for the IEA Reading Literacy Study, Language Education Study, CIVED, PIRLS 2001, and TIMSS 1995, 1999, 2003, and 2007. Dr. Papanastasiou is the author or editor of more than ten books and many articles related to IEA studies. He founded the IEA International Research Conference and organized the first conference in Cyprus in 2004, and was awarded the title of Honorary Member of the IEA in the 2007 General Assembly meeting in Hong Kong

*Address:* 4, Chr. Kannaouros St. Latsia 2220, Nicosia, Cyprus

*Email:* edpapan@ucy.ac.cy

---

Co-author of chapter 9, "IEA International Research Conferences".

**Papanastasiou, Elena C.** is Associate Dean and Associate Professor at the



University of Nicosia, Cyprus, where she also teaches courses in Educational Research, Measurement and Quantitative Methods. She received her Honors BSc in Elementary Education from The Pennsylvania State University, and her PhD in Measurement and Quantitative Methods from Michigan State University. Since 2001, Dr. Papanastasiou has also worked at the University of Kansas and the University of Cyprus. Dr. Papanastasiou is mostly interested in the areas of Assessment and Attitudes toward Research. So far, she has published the Attitudes Toward Research Scale (2005) that is widely used in various countries, in addition to a book on educational research (in Greek) and more than 30 peer-reviewed articles. In 2011, she also became a General Assembly member of the IEA.

*Address:* University of Nicosia, Department of Education, 46 Makedonitissas Av., 1700 Nicosia, Cyprus

*Email:* Papanastasiou.e@unic.ac.cy

---

Co-author of chapter 9, "IEA International Research Conferences".

**Pelgrum, Willem J. (Hans)** has been the international coordinator of the



IEA Computers in Education studies and of SITES Module 1 and SITES 2006. He worked as researcher from 1978 to 2005 at the University of Twente, Enschede, The Netherlands, and owns at present a consultancy firm EdAsMo (Educational Assessment, Monitoring and Evaluation). He has run several projects for the European Commission in the areas of indicators and ICT. He is involved as expert in the selection of proposals for several research programs from the European Commission.

*Address:* EdAsMo, Brandijzerhoek 2, 7546LK Enschede;

*Email:* w.pelgrum@inter.nl.net

Co-author of chapter 6, "IEA assessments of Information and Communication Technologies (ICT) in education" and chapter 19, "IEA Experiences from Latvia, Lithuania, Slovak Republic and Slovenia".

**Plomp, Tjeerd**, MSc in mathematics and PhD in social sciences (both Free



University Amsterdam), professor emeritus of education (educational technology and curriculum) at the University of Twente in Enschede, The Netherlands. He represented The Netherlands in the General Assembly of the International Association for the Evaluation of Educational Achievement (IEA) from 1987 - 1989, and was chair of the IEA from 1989 to 1999. During his time with IEA, he served as Chair of the International Steering Committee of the Computers in Education Study (COMPED), the Third International Mathematics and Science Study (TIMSS), and the Second International Technology in Education Study (SITES). He was also study director for IEA SITES 2006, which surveyed schools and (mathematics and science) teachers on ICT use and ICT-related pedagogical approaches.

*Address (postal):* Karel van Gelderlaan 34, 6861CC Oosterbeek, The Netherlands

*Email:* t.plomp@utwente.nl

Author of chapter 2, "IEA from 1988 - 1994: An Organization in Transition"; co-author of chapter 6, "IEA Assessments of Information and Communication Technologies (ICT) in education"; and roundtable participant in chapter 21, "Roundtable at the 49th General Assembly, Berlin 2008".

**Postlethwaite, T. Neville** (1933–2009) was, from 1976 until his retirement, professor of comparative education at the University of Hamburg in Germany. He began his professional career as Lecturer at St. Albans College of Further Education and then served as research officer at the National Foundation for Educational Research (1961–1963), which was followed by an international career. Postlethwaite was, for example, the first Executive Director of IEA (1962–1972), and also staff member of the UNESCO International Institute for Educational Planning in Paris (1972–1976). In 1978, he succeeded Torsten Husen as Chair of IEA (1978–1986), with whom he also was co-editor-in-chief of the first and second editions of the *International Encyclopedia of Education*. Postlethwaite was international coordinator for a number of IEA studies, such as the First International Mathematics Study (FIMS), Six Subject Survey, and Reading Literacy Study. He is author or co-author of a number of books, research reports, journal articles, and special consultancy reports. He was also Honorary Member of the IEA.



Roundtable participant in chapter 21, "Roundtable at the 49th General Assembly, Berlin 2008".

**Purves, Alan C.** (1931–1996), Professor of English and later Professor of Education and Head of the Laboratory School at the University of Illinois (1970–1986). Professor of Education and Acting Director of the Center for Writing and Literacy, State University of New York at Albany (1986–1996). Chairman of IEA (1986–1990). His work in the 1960s and 1970s on response to literature had a major impact on the teaching of literature in the classroom. He was a widely recognized expert on children's literature. Alan's contributions to the National Council for Teachers of English (NCTE) were without parallel: among several tasks, he was editor of "Research in the Teaching of English" (RTE) for 1971–1977, President of the NCTE 1979–1980, author of many NCTE books and contributed scores of articles to its journals. His work covered research in teaching and testing in literature and writing, curriculum and educational policy, educational objectives in the humanities and literacy, and literacy criticism. He continued to explore new avenues to the end of his life, for instance, the use of portfolios, and to reflect on relationships between society, technology and literacy and extending this further to the relationship between religion and literacy. Alan was widely known in language education circles the world over. He was International Coordinator of the IEA Literature Education and Written Composition Studies (1970–1971 and 1985).



Author of chapter 23, "The Evolution of the IEA: A Memoir"

**Robitaille, David** is professor emeritus of mathematics education at the



University of British Columbia in Vancouver, Canada. He first became associated with IEA in 1978 as the General Assembly representative for British Columbia and as NRC for the province in SIMS. He co-edited one of the three volumes emanating from that study. Later he served as a member of the Standing Committee of IEA and was the International Coordinator for TIMSS until 1991. Since 2001 he has served as chairperson of IEA's Publication and Editorial Committee and of its Awards Committee. Dr. Robitaille began his career as a teacher in the public school system in Montreal. After obtaining his PhD at Ohio State University, he accepted a position at UBC. He is the author of more than 100 papers, reports, books, and monographs. His research has been supported by grants from numerous Canadian and international agencies.

*Address:* Department of Curriculum and Pedagogy, Faculty of Education, 2125 Main Mall, University of British Columbia, Vancouver, B.C., Canada V6T 1Z4

*Email:* david.robitaille@ubc.ca

---

Roundtable participant in chapter 21, "Roundtable at the 49th General Assembly, Berlin 2008".

**Schulz, Wolfram**, Dipl. Pol. and PhD in Political Science (Free University



Berlin and University of Rostock), Research Director International Surveys at the Australian Council for Educational Research (ACER) in Melbourne. Prior to joining ACER, Dr Schulz worked for three years as Associate International Co-ordinator of the IEA Civic Education Study at the Humboldt University Berlin. During his time at ACER, he was Research Director of the International Study Centre of the IEA International Civic and Citizenship Education Study (ICCS 2009), which assessed learning outcomes of civic and citizenship education in 38 countries in 2008 and 2009.

*Address:* Australian Council for Educational Research, 19 Prospect Hill Rd, Camberwell VIC 3124

*Email:* schulz@acer.edu.au

---

Co-author of chapter 8, "The International Civic and Citizenship Education Study (ICCS)".

**Schwille, John (Jack)** (BA, Harvard University; Ph.D, University of



Chicago). Professor, 1982-present, and Assistant Dean for International Studies in Education, 1984-present, College of Education, Michigan State University. Active in IEA since 1972: Spencer Fellow and Research Officer, IEA Secretariat at University of Stockholm, contributor to First IEA Study of Civic Education and organizer of databank for IEA Six Subject Survey, 1972-74; part of international leadership team for Second International Mathematics Study (SIMS) while at the National Institute of Education, Washington DC, and at Michigan State University, 1976-84; national and international panels, 1984-1996, advising IEA, including Board on International Comparative Studies in Education (BICSE) at the National Academy of Sciences ,1990-96; member of steering committee, IEA Second International Study of Civic Education (CIVED), 1994-2001; co-director of first IEA Teacher Education and Development Study (TEDS-M); Honorary Member of IEA, 2010-present.

*Address:* College of Education, Michigan State University, 517 Erickson Hall, East Lansing, Michigan, 48823 USA *Email:* jschwill@msu.edu

Co-author of chapter 4, "The IEA Studies in Civic Education: Their History from the Late 1960s to 2005"; and author of chapter 27, "Experiencing Innovation and Capacity Building in IEA Research - A Memoir, 1963-2008".

**Suter, Larry E,** BA 1965, Sociology, College of Idaho; MA,1968, PhD 1975, Sociology Duke University). Retired from the National



Science Foundation, where he was program director for research on learning (1990 to 2011). He created the program of studies at the National Center for Education Statistics to support the development of data bases in the Second International Mathematics Study (1981 to 1986), the Second International Science Study (1986), the Computers in Education Study (1988), the International Reading Study (1989), and the initial phases of the Third International

Mathematics and Science Study. From the National Science Foundation in 1990 he continued as program director for TIMSS in the early development stages and helped manage the continuation of the study until 2003. He also was the program manager for the National Academy of Sciences Board on International Comparative Studies in Education from 1990 to 2003. He was one of the program directors that supported the Teacher Education Study. He met Neville Postlethwaite and Torsten Husen in 1982 to begin negotiations on how to create an international data set that could be analyzed by statisticians and researchers.

*Address:* 2515 Gadsby Place, Alexandria, VA 22311 *Email:* Larrysuter@comcast.net

Author of chapter 29, "Guiding International Comparative Studies (1981 to 2009): Reflections of a U.S. Program Director".



**Takala, Sauli**, PhD in 1984 at the University of Illinois at Urbana-Champaign, Professor emeritus in Applied Linguistics, University of Jyväskylä. He has extensive experience in research on language testing and assessment, in language policy and planning, curriculum development, and teachers' in-service education. He has published widely in Finnish, Scandinavian and international journals, and supervised and reviewed a great number of doctoral dissertations in Finland and other Nordic countries.



He co-ordinated the IEA International Study of Writing and helped to plan and co-ordinate the EU-funded internet-based DIALANG diagnostic assessment project. He has had close links with the Council of Europe's work on modern languages. He has been editor for the Finnish journal of educational research, a co-editor of the Scandinavian Journal of Educational Research, a member of the editorial board of Language Testing. He served as the second President of EALTA (European Association for Language Testing and Assessment).

*Address:* Centre for Applied Language Studies, University of Jyväskylä, Finland

*Email:* shtakala@hotmail.com

Author of chapter 5, "The International Study of Writing".

**Thomson, Sue** is the Director of Educational Monitoring and Research at the Australian Council for Educational Research. After gaining a Master of Educational Studies and a PhD in Education from Monash University, Dr Thomson joined ACER in 1998, working on the Longitudinal Surveys of Australian Youth (LSAY) project, as well as on projects examining numeracy over the transition from pre-school to school and science teaching internationally. She has been the National Research Coordinator for Australia for TIMSS since 2001 and has managed Australia's participation in all cycles of TIMSS since then as well as Australia's first participation in PIRLS in 2010. Dr Thomson has co-authored national reports for Australia on each cycle of TIMSS, as well as a number of other more in-depth analyses which have been published by ACER or have been presented to the IEA bi-annual conference.



*Address:* ACER, 19 Prospect Hill Rd, Private Bag 55, Camberwell, Victoria 3124, Australia

*Email:* thomsons@acer.edu.au

Co-author of chapter 14, "The Contribution of IEA Research Studies to Australian Education".

**Tiana Ferrer, Alejandro**, Professor of Theory and History of Education at



Spain's National Distance Teaching University (UNED) and since 2008 Director General of the Centre for Advanced Studies of the Ibero-American States Organization for Education, Science and Culture (OEI). As Director of the Centre for Educational Research and Documentation, and subsequently as Director of the National Institute for Quality of Education of the Ministry of Education, he was Spain's representative to the IEA (1989-1996), a member of the IEA Standing Committee (1992-1994) and Chair of IEA (1999-2004). He served as Secretary General of Education for the Ministry of Education and Science of Spain (2004-2008). Author or joint author of nineteen books and over 150 studies on different subjects related to education.

*Address:* Universidad Nacional de Educación a Distancia, Facultad de Educación, Paseo Senda del Rey, 7, 28040 Madrid (Spain).

*Email:* atiana@edu.uned.es

Author of chapter 24, "Experiencing the World as an Educational Laboratory"; roundtable participant in chapter 21, "Roundtable at the 49th General Assembly, Berlin 2008".

**Torney-Purta, Judith**, B.A., Psychology, Stanford University; Ph.D, Human



Development (Social Sciences), University of Chicago; Professor of Psychology, the University of Illinois Chicago (1968-1981); Professor of Human Development, the University of Maryland, College Park (USA) since 1981. She received the American Psychological Association's Award for Distinguished Contributions to Psychology Internationally and the American Orthopsychiatric Association's Langer Award for Social Advocacy and Human Rights. At the U.S. National Academy of Sciences she was a member of the Board

on International Comparative Studies in Education and is currently a member of the Committee for Psychological Science. She was senior author of the book reporting the first IEA civic education study (1975). She led phases 1 and 2 of the second civic education study (authoring or editing four books). She and her students have conducted and published extensive secondary analysis of CIVED data. A recent publication: Handbook of Research on Civic Engagement in Youth.

*Address:* Department of Human Development, University of Maryland, College Park, MD 20742 USA

*Email:* jtputa@umd.edu

Co-author of chapter 4, "The IEA Studies in Civic Education: Their History from the Late 1960s to 2005"; author of chapter 25, "50 Years of IEA: Some Reflections about its History"; and editor of chapter 21, "Roundtable at the 49th General Assembly, Berlin 2008".

**Travers, Kenneth J.**, MEd in mathematics education (University of British



Columbia); PhD in mathematics education (University of Illinois at Urbana-Champaign); professor emeritus of mathematics education at the University of Illinois at Urbana-Champaign. He was chairman of the International Mathematics Committee for the Second IEA International Study of Mathematics, 1976-1987.

Travers was also Director of the U. S. Coordinating Center for SIMS.

*Address:* Office for Mathematics, Science and Technology Education, University of Illinois, Suite 102, 505 E. Green Street, Champaign,, IL 61821 USA.

*Email:* ktravers@illinois.edu

Author of chapter 3, "The Second International Mathematics Study (SIMS): Intention, Implementation, Attainment".

**Wagemaker, Hans** is the executive director of the International



Association for the Evaluation of Educational Achievement, a position he has held since 1997. He holds BA (Hons.) and MA (Hons.) degrees in education from the University of Otago, New Zealand and was awarded a PhD from the University of Illinois in 1981. He returned to New Zealand in 1982 to work for the Ministry of Education. As Senior Manager Research and International he led New Zealand's delegations to the APEC and represented New Zealand's interests in UNESCO and the OECD. In 1989 he was awarded a Nuffield Commonwealth Fellowship and in 1996 was given the McKenzie Award for outstanding contribution to research in education by the New Zealand Association for Research in Education. In 2009 he was awarded the University of Illinois, College of Education, Distinguished Alumni Award. As executive director of the IEA he is responsible for the management of IEA's international research projects, a Secretariat in Amsterdam and a large data processing and research center in Hamburg, Germany.

Hans has consulted for the Inter-American Development Bank and UNESCO and has worked extensively with the World Bank to advance a common interest improving educational systems in developing countries.

*Address:* IEA Secretariat, Herengracht 487, 1017 BT Amsterdam, The Netherlands

*Email:* hanswagemaker@compuserve.com

Author of chapter 11, "IEA: International Studies, Impact and Transition".

**Wolf, Richard M. (Dick)** (1934-2004) worked from 1968-1999 at Teachers



College, Columbia University, as Professor of Psychology and Education. He was Chair of the Department of Measurement, Evaluation and Statistics. His research focused on measurement and evaluation in education. He participated in the IEA First Mathematics Study (FIMS, 1964), the Six Subject Survey (1970-1971), the Second International Science Study (SISS, 1983-1986), and the Computers in Education Study (COMPED, 1989, 1992). He

was for many years the Chair of the IEA Publications and Editorial Committee. Dick Wolf was a member of the International Academy of Education, and authored more than 50 publications; he was also consultant to various national and international organizations in education.

---

Author of chapter 12, "The Contribution of IEA to Research and Education".





