



# Pitfalls of Using International Test Scores to Inform Policy

1. Three Common Misuses
2. How They Mislead

Tom Loveless

Cape Town, South Africa, June 22, 2015

## ➤ Three Common Misuses

- The Problem of Rankings
  - Focusing on A+ Countries
  - Dubious Causal Inferences
-





## EDUCATION IN JAPAN COMMUNITY Blog

*Features & info on educating kids in Japan*

 [Subscribe to feed](#)

[Home](#) [ABOUT US](#) [CULLING COLLEGE INFO](#) [THE SCOOP ON SCHOOLS](#)  
[HEADING FOR HIGH SCHOOL](#) [PARENTING POTPOURRI](#) [OF METHODS](#)  
[& PHILOSOPHIES](#) [EDUCATION IN GENERAL, ISSUES & OPTIONS](#) [KIDS](#)  
[HEALTH & SAFETY ISSUES](#) [FIELD TRIPPER'S CORNER](#) [STUDENT LIFE](#)  
[IN JAPAN](#) [THE BOOK NOOK](#) [RESOURCE ROOM](#) [HOMESCHOOLING](#)  
[/ AFTERSCHOOLING](#) [EDU NEWS](#) [BULLETIN BOARD](#) [A SITE MAP](#)  
[References](#)

### BLOGROLL

[A Resource on Tohoku-Kanto Earthquake-Tsunami-Nuclear Disaster](#)

[Blog away and grow a tree!](#)

[Directory of International Schools & internationally-oriented schools](#)

[Directory of Major Japanese Language Schools](#)

[Directory of schools that offer bilingual educational programmes \(from the Education in Japan website: our portal site\)](#)

[Donate to Tohoku-Kanto Quake/Tsunami Victims Without Being Scammed](#)

[Education in Japan \(in a nutshell\)](#)

[Education in Japan \(Our portal website\)](#)

[Examination for Japanese University Admission for International Students \(IUJ\)](#)

[Financial Aid and Scholarships for Japanese or domestic students](#)

[Free Brain Games at](#)

## Edu Watch: Japanese students improve in PISA rankings & other news;

December 11, 2010 in [1](#)

Hello to readers old and new of the EIJ blog,

The spotlight on educational news this month is dominated by introspection by various educational agencies over the results from 2010 PISA study reports.

Here's our rundown on the news on the 1. educational scene:

[Academic rankings bring tempered praise / Japanese students fare better in latest international tests, but alarm over declining standards persists \(Dec.9, 2010\)](#)

"Japanese students in 2009 showed improved reading, math and science skills in international academic aptitude tests compared to three years earlier, but were outdone by their counterparts in Shanghai who ranked top in all three fields, the Organization for Economic Cooperation and Development reported Tuesday...

Japan's rankings in the 2003 and 2006 PISA tests for 15-year-old students caused wide public concern about a general decline in the academic ability of Japanese children.

In 2009, however, Japan's rank in reading rose to eighth among the 65 countries and regions involved in the program, which saw 470,000 students take the exams. Japan was ranked 15th in reading in 2006.

The nation's students were ranked fifth in applied scientific skills, up from sixth in 2006, and ninth in applied mathematical skills, rising from 10th.

Asian countries and regions featured in the upper rankings in all three skill sets, with Shanghai claiming top position in all of them, the results of the triennial tests show.

“The nation’s students were ranked fifth in applied scientific skills, up from sixth in 2006, and ninth in applied mathematical skills, rising from 10<sup>th</sup>.”

This is misleading.

---



**Exhibit 1.2 TIMSS 2007 Multiple Comparisons of Average Mathematics Achievement**

Instructions: Read across the row for a country to compare performance with the countries listed along the top. If the average achievement of the country in the row is significantly lower than that of the comparison country, or if there is no statistically significant difference between the average achievement of the two countries, the cell is shaded gray.

[illegible]

Exhibit 1.3 Trends in Mathematics Achievement – 1995 Through 2007

TIMSS2007  
Mathematics 4<sup>th</sup> Grade

Country		Average Scale Score	2003 to 2007 Difference	1995 to 2007 Difference	Mathematics Achievement Distribution				
Hong Kong SAR									
†	2007	607 (3.6)							
	2003	575 (3.2)	32 (4.8) ▲						
	1995	557 (4.0)		50 (5.4) ▲					
Singapore									
	2007	599 (3.7)							
	2003	594 (5.6)	5 (6.7)						
	1995	590 (4.5)		9 (5.9)					
Chinese Taipei									
	2007	576 (1.7)							
	2003	564 (1.8)	12 (2.5) ▲						
Japan									
	2007	568 (2.1)							
	2003	565 (1.6)	4 (2.6)						
	1995	567 (1.9)		1 (2.8)					

Trends in International Mathematics and Science Study (TIMSS) 2007



## Lesson: Don't Misuse Rankings

1. The difference between two close rankings may not be statistically significant. Check the “Multiple Comparisons” tables.
  2. A change in ranking from one test to the next may not be statistically significant. Check the “Trends” tables.
  3. Rankings are not equal interval. A 19 point gain may move a nation up one ranking, several rankings....or no ranking at all.
-

## **Example #2:**

### ➤ The Problem of “A+” Countries

Pointing to a handful of top scoring countries and saying, “They are doing Policy X, therefore Policy X is good.”

---



**TIMSS Eighth Grade Math Scores for “A+” Countries, 1995–2011**  
**(Ranked by TIMSS 1995 Score)**

Table

**1-3**

Nation	TIMSS 1995	Last TIMSS (Year)	Change
Singapore	609	611 (2011)	+2
Japan	581	570 (2011)	-11*
Korea	581	613 (2011)	+32*
Hong Kong	569	586 (2011)	+17*
Belgium (Flemish)	550	537 (2003)	-13*
Czech Republic	546	504 (2007)	-42*

\*  $p < .05$

Note: The U.S. had a score of 492 in 1995 and 509 in 2011, a change of +17\*

Source: TIMSS 2011 International report

1. Scores change. A+ countries may decline. If so, would you want your country to follow their policies?
  2. Should look across the entire distribution of scores. For example, imagine that Policy X is a policy adopted in all of the A+ countries. Should one assume that Policy X contributes to high TIMSS scores? What if Policy X is also found in countries at the bottom of the distribution—among low scoring countries? Should one then assume that Policy X contributed to those countries' low scores on TIMSS?
  3. Statisticians call this “selecting on the dependent variable,” which often misleads.
-



# **Example #3: What Caused Poland's PISA Gains?**

**It's A Mystery....**



What is the most popular explanation for Poland's gains?








From the 2006 PISA Executive Summary:

“A long-term trend in OECD countries has been to reduce the amount of separation and tracking in secondary education. The most recent major example of this is Poland, whose reading results before and after this education reform are reported in PISA. “

Executive Summary, *PISA 2006: Science Competencies for Tomorrow's World*, page 39.

---



“Here, [Poland] an improvement in results among lower ability students immediately after the reform was not at the expense of higher ability students, whose results also rose in the subsequent period.”

Executive Summary, *PISA 2006: Science  
Competencies for Tomorrow's World*, page 39.

---




## From the World Bank:

“Poland’s reading score was below the OECD average in 2000, at the OECD average in 2003, and above the OECD average in 2006, ranking 9th among all countries in the world....

“With regard to the factors responsible for the improvement, the delayed tracking into vocational streams appears to be the most critical factor.”

Successful Education Reform: Lessons from Poland, 2010,  
*Europe and Central Asia Knowledge Brief* (Nov. 2010,  
Volume 34), page 3.

---



By 2009, the prevailing story was that tracking reform, more than anything else, had boosted Poland's reading scores.

In 1999, Poland had extended compulsory education from age 15 to age 16 and created a new lower secondary school that delayed tracking into vocational programs.

---




# Table of Contents

## Chapter on Poland

### *From Strong Performers and Successful Reformers in Education: Lessons from PISA for the United States*

#### **Poland: Secondary education reform.....224**

- *A highly tracked education system pre-1989*
- *Education reforms since 1989: The birth of the technical lyceum*
- *Structural reforms of the late 1990s*
- *The results: A remarkable turnaround*



So did tracking or streaming  
reform have anything to do  
with Poland's success?

Answer is:

Maybe. But maybe not. We  
really don't know.

---





One of the strongest pieces of evidence:

- The significant gains made by Poland's low performing students.

These are students who would have been assigned to the vocational track and now, presumably, were benefitting from an additional year of exposure to an academic curriculum.

---

# But after the 2009 PISA we learned.....

“In nearly all the countries that showed improved performance during the period, [2000-2009] the percentage of low performers dropped, meaning that the number of students who scored below the PISA baseline reading proficiency Level 2 was significantly smaller in 2009 than in 2000.”

(From: “*Improving Performance: Leading from the Bottom*,” OECD, 2011, page 1.)

---



Quote continues:

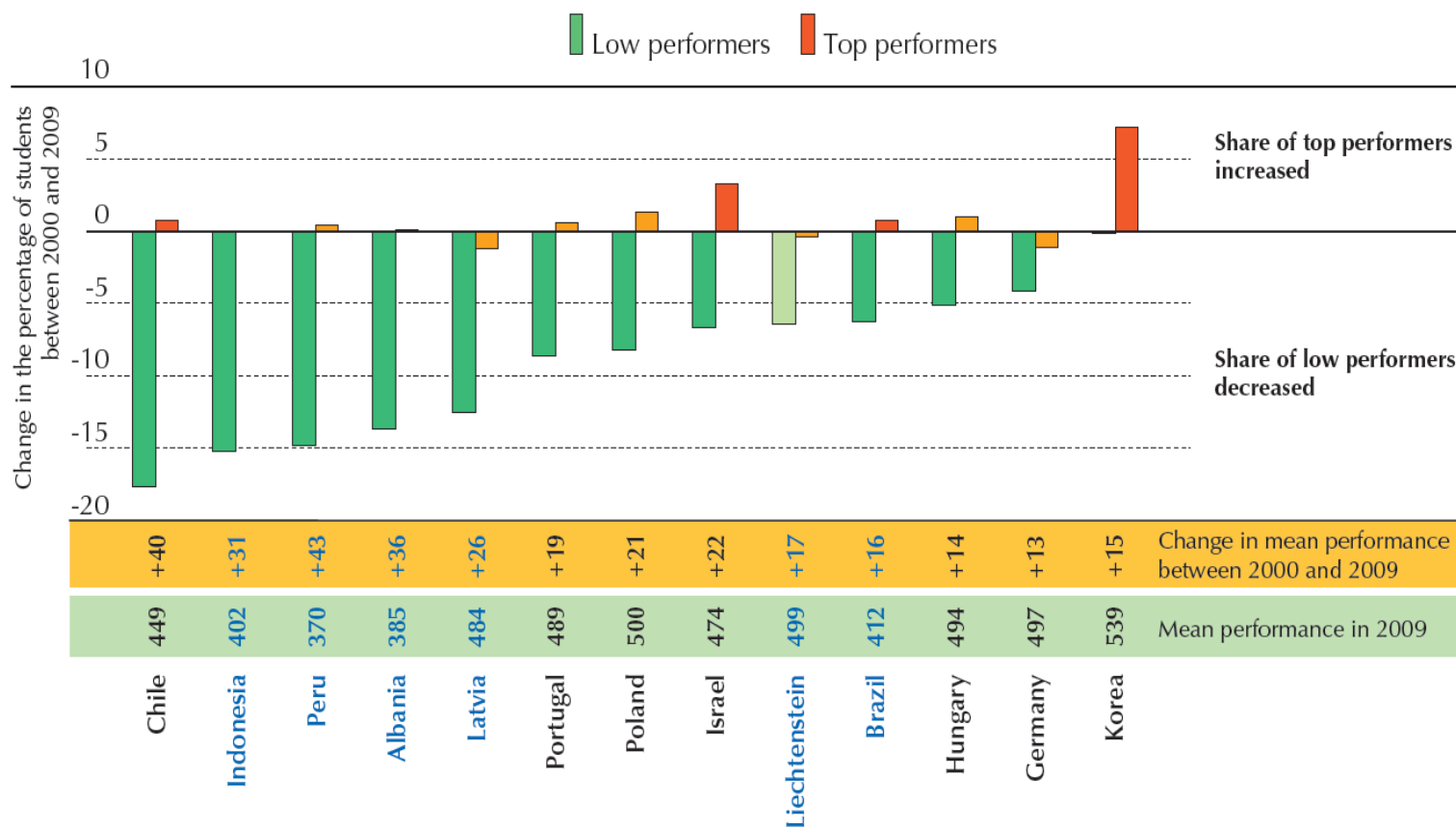
“While the percentage of low performers changed only slightly, on average across OECD countries, it dropped from nearly half (48%) of all 15-year-old students to below one-third (31%) [in the countries with improving PISA scores].”

So Poland was not alone.

(From: “*Improving Performance: Leading from the Bottom*,”  
*OECD, 2011, page 1.*)

---

## Change in the percentage of low and top performers in reading since 2000



Note: Values that are statistically significant are marked in a darker tone.

Source: OECD, PISA 2009 Database, Tables V.2.1 and V.2.2.



## Doubt Creeps into the Story

- True, low achieving students in Poland made big gains in reading.
  - But low achievers made similar gains in all 13 countries with growth from 2000 to 2009. The other 12 countries didn't reduce tracking, and many made even larger gains in reading than Poland.
-

# Poland's 1999 Reforms

## Much More Than Tracking

Key elements:

- Decentralization
- Greater Autonomy for schools
- Increase in Teacher salaries
- New system of national assessment
- Adoption of a core curriculum and national standards
- Teacher education reform at university level
- New system of teacher promotion



Another possible factor:

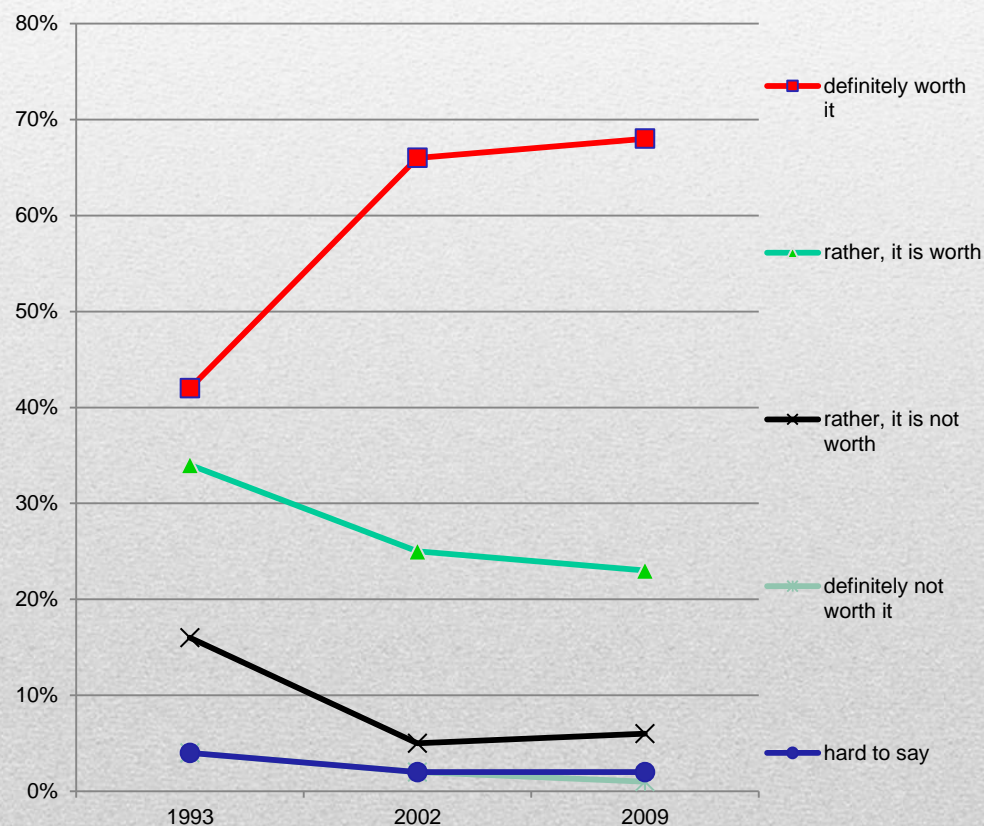
➤ Public attitudes towards education were shifting.

---

# Context of the comprehensive reform – educational aspiration

Significant growth of citizens' educational aspirations in 1993 - 2009:  
91% of adult Poles believe that it is important to get an education, of which 68% are strong supporters of this idea

Is it worth gaining education?  
1993 – 2009





# Did Tracking Reform Lead to Poland's Gains on PISA in Reading from 2000 to 2009?

Answer: Maybe, but there are reasons for doubt

1. Other countries show gains among low performers, too, without tracking reform.
  2. Besides tracking reform, several big reforms were put in place at the same time. Difficult to tease out the effects of one policy.
  3. Attitudes towards education were changing at the same time.
-

# Lessons for Policy Analysts and Policymakers

1. It is not a good idea to single out one policy as causing test score changes when several policies are being implemented at the same time.
  2. It is not a good idea to single out one country. Others may be trying the same policies and failing or experiencing even greater success while trying other policies.
  3. Singling out one policy in one country as a model combines the problems of #1 and #2. It is a doubly bad idea.
-