Analysis of ILSA data: Comments on Gustafsson and Chmielewski & Dhuey

Daniel Koretz
Harvard Graduate School of Education

National Academy of Education
Workshop Series on International Large-Scale Assessment
Workshop I: Directions for Improving ILSA Design and Analysis
Washington, DC
June 17, 2016
“You can’t fix by analysis what you’ve bungled by design.”
Light, Singer, & Willett, 1990

Today: “You can’t entirely compensate by analysis for what is unavoidably limited by design”
Some threats to robust inference with ILSAs

1. Inconsistencies between tests across countries (interactions between test design and curricula)
2. Differences among ILSAs in scaling and conditioning
3. Differences in sample design and nonresponse
4. Weak set of predictors
   a) Limited in scope and source
   b) Survey variables may behave differently across countries
   c) Direct measures may vary in meaning across countries
   d) Proxies rely on untested similarities in relationships with causal factors
   e) Cross-national comparisons create additional omitted variables
2003 TIMSS & PISA math

$r = .84$
<table>
<thead>
<tr>
<th>Correlation Details</th>
<th>Correlation Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>TIMSS vs. PISA, all</td>
<td>0.84</td>
</tr>
<tr>
<td>TIMSS vs. PISA, -2</td>
<td>0.67</td>
</tr>
<tr>
<td>CUNY, Regents vs. SAT math, student</td>
<td>0.77</td>
</tr>
<tr>
<td>CUNY, Regents vs. SAT math, school</td>
<td>0.86</td>
</tr>
<tr>
<td>ITBS G8, reading vs. math, student</td>
<td>0.73</td>
</tr>
<tr>
<td>ITBS G8, reading vs. math, school</td>
<td>0.88</td>
</tr>
</tbody>
</table>
TIMSS/PISA: Korea-US

**PISA Math 2003 (15-Year-Olds)**

**Korea**

0.62

Standard Deviations

**United States**

**How Far is the U.S. from Korea?**

How much better do students in Korea perform than those in the United States?

**TIMSS = PISA**

Mean = 500 & Std. Dev. = 100

**TIMSS Math 2003 (Grade 8)**

**Korea**

1.06

Standard Deviations

**United States**

**Note:** Distances are expressed in standard deviation units for the United States.
Math SDs, 2003, PISA & TIMSS grade 8

\[ r = 0.07 \]
TIMSS G8 math SDs, 24 countries, 95 & 99

![Graph showing TIMSS G8 math SDs for 24 countries in 1995 and 1999. The graph plots 1995 SD on the x-axis and 1999 SD on the y-axis. Points represent each country, with South Africa and the US marked. There is a downward trend line indicating a decrease in SDs over the years.]
Variation in age (synthetic cohorts)

Relative inequality increases in every country with tracking except the Slovak Republic, while relative inequality decreases in every country without tracking except for Sweden and Latvia.

Fig. 1. Inequality in Primary and Secondary School
Notes: Standard deviation of test scores in the national population (difference from international average of national standard deviations in each test). Countries with a tracked school system before the age of 16 have solid lines, countries without tracking before age 16 have dashed lines.

Source: Hanushek & Woessmann 2006
PISA 2003: robustness across strands:
top 10 performers by strand
Key RHS variables in Gustafsson

- ‘Parental support for learning’
- Amount of homework
TIMSS 1995, “Mother thinks it’s important for me to do well in math,” BSBMMIP2
Conclusions

- ILSAs offer advantages, but also necessarily entail both design limitations and compromises.

- Appropriate use requires evaluating each use (causal or descriptive) against limitations as well as opportunities.

- Risks include:
  - (Unrecognized) failure to replicate
  - Conflating differences among surveys with predictors
  - Misinterpretation of RHS variables

- Avoid analysis that “is beyond the carrying capacity of the data” (H. Braun)
Supplementary slides
2003 TIMSS & PISA math, excl. Indonesia, Tunisia

$r = .67$
### TIMSS 2011 G8 & PISA 2012

**Math Subscale Correlations**

<table>
<thead>
<tr>
<th></th>
<th>TIMSS 2011 G8</th>
<th>PISA 2012</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum</td>
<td>0.87 data, algebra</td>
<td>0.90 uncertainty, space</td>
</tr>
<tr>
<td>Maximum</td>
<td>0.97 number, geometry</td>
<td>0.96 uncertainty, change</td>
</tr>
<tr>
<td>Mean</td>
<td>0.95</td>
<td>0.94</td>
</tr>
</tbody>
</table>

(16)
TIMSS 2007 grade 8: lack of robustness across parts
US Math trends, NAEP vs. TIMSS

Grade 4

Grade 8

SDs


SDs


timss  naep

(18)
TIMSS G8 SDs, 2007 & 2011

$r = 0.80$
TIMSS 1995, maternal education