



# International Educational Assessments: Cautions, Conundrums, and Common Sense

National Academy of Education

Tom Loveless

April 6, 2018



“The research questions we want to answer and the research questions we can answer are often not the same.”

*International Education Assessments, page 68.*



The research questions the public wants answered are:

1. Why did we [fill in name \_\_\_\_\_ of Country A] score so badly?
2. Why did they [fill in name \_\_\_\_\_ of Country B] score so much better?
3. What are they in Country B doing that we in Country A should be doing?



TABLE 5-1 Categories of Analytical Strategies Used by Researchers

Category	Strategy
1	Analyze policy variation across countries.
2	Analyze policy variation within countries.
3	Analyze repeated cross-sectional ILSA data to look at variation across birth cohorts or generations of students.
4	Analyze repeated cross-sectional ILSA data to look at variation across age within the same birth cohort within countries; known as “synthetic cohorts.”
5	Analyze rare truly longitudinal ILSA data that follow the same students over time.

SOURCE: Chmielewski & Dhuey, 2017.



# Limitations of International Assessments

Sean Reardon of Stanford University. P.19

“Test scores are the product of the full set of experiences and opportunities a kid has to learn in [his or her] entire life, some of which happens in school, but a lot of which happens outside of school, in the home, after school, in preschool.”

Henry Levin of Teachers College p. 23.

“Levin stated, ‘these studies can only be correlational. They are not causal studies. They are not longitudinal [at the student level]. We need to keep that in mind because everyone here knows that. But when you go to the PISA results and the league tables, they are interpreted as being causal inferences drawn from certain countries.’”



## A Conundrum

The public, press, and policy makers want ILSAs to explain the causes of low and high achievement—and, in the case of low achievement, to identify policy solutions--even though the assessments are not able to do so.



## Then What Good Are ILSAs ?

- Monitoring national achievement is worthy by itself. Most homes have a thermometer in the medicine cabinet. It can't diagnose (identify causes) but it can raise an alarm if something is wrong.
- Hypothesis building—alert policy analysts to promising topics needing more rigorous causal designs (RCT or quasi-experimental design).



## Two Ideas

Idea #1 The organizations that administer ILSAs and release results would be wise to devote greater resources to preparing reporters and providing more guidance on what can and cannot be inferred from results. (p. 70)

--Worth a try.

--But New Media doesn't follow the old rules.



## Two Ideas (cont.)

Idea #2 An impartial, national board could be created and charged with providing ongoing guidance on ILSA design, analysis, reporting, and interpretation. (p.70)

- I'm not in favor of guidance on design, analysis, or reporting. Too many cooks in the kitchen.
- International body to issue a document laying out a set of Principles for Interpreting Evidence from International Assessments.
- The Principles must have some bite. Creation of a standing committee of international scholars to review reports and call out gross violations of the principles.



# A Final Idea on Governance

- Researchers must have a more prominent role in governing OECD-PISA and IEA.
- Not simply an advisory role, but making governing decisions.
- International Assessments began with IEA, founded and governed by education researchers. Today, government officials dominate the governance of both IEA and OECD-PISA.
- Consequence: Those who create education policy and are held accountable for policy's success wield power over international efforts to gather information on national systems' performance. The conflict of interests is obvious.