

International Educational Assessments: Cautions, Conundrums and Common Sense

Summary of Recommendations

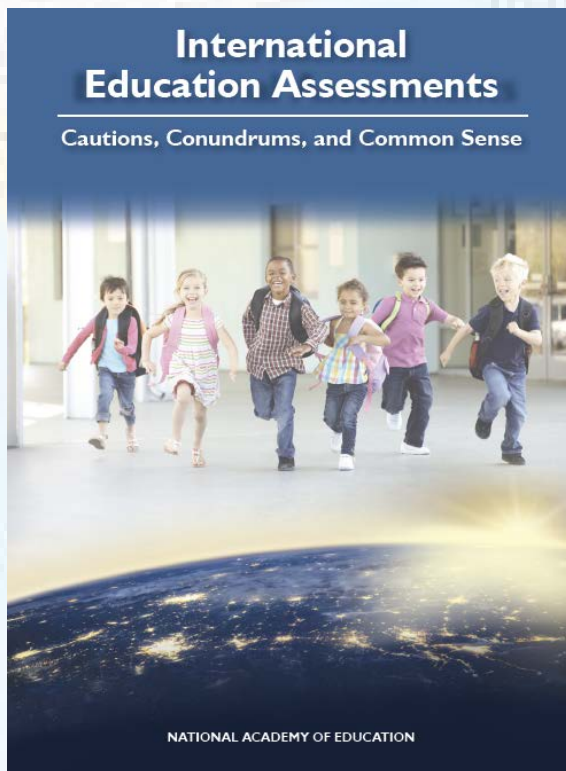
Judith D. Singer, PhD

James Bryant Conant Professor of Education
Senior Vice Provost for Faculty Development and Diversity
Harvard University

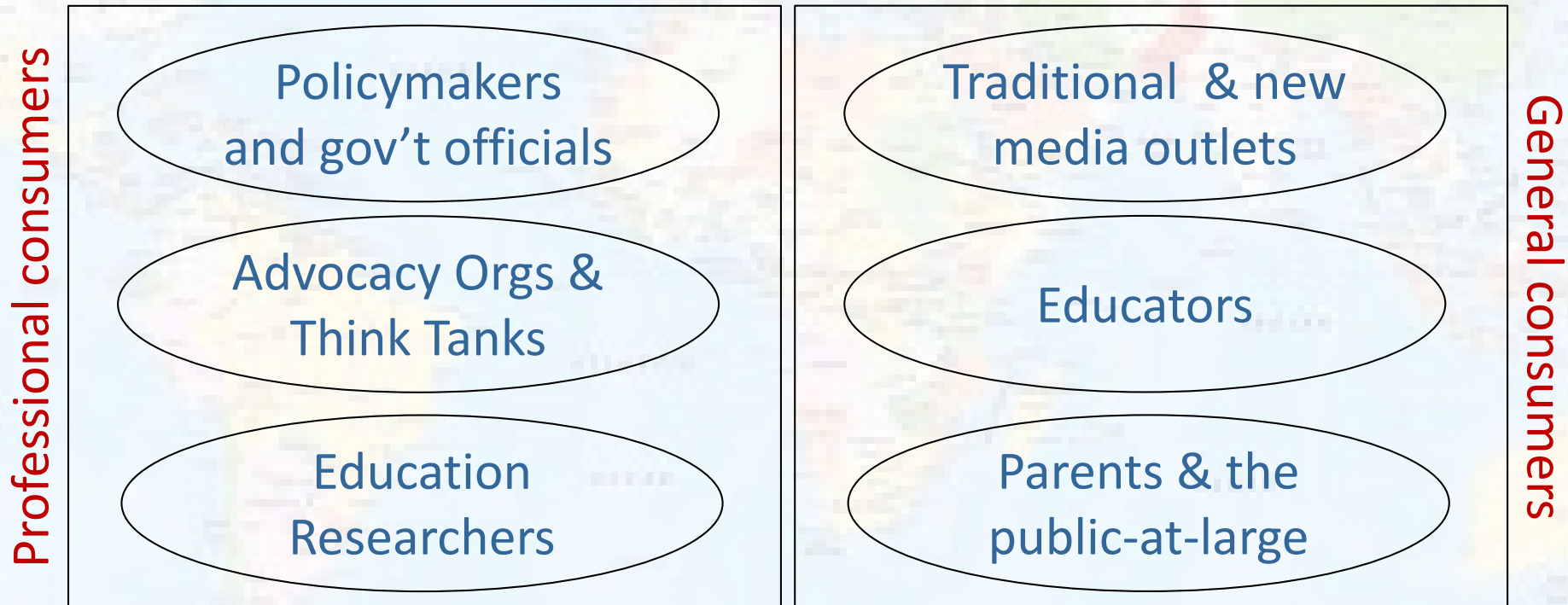
Henry I. Braun, PhD

Boisi Professor of Education and Public Policy
Director, Center for the Study of Testing, Evaluation, and Education Policy
Boston College

April 2018



The ILSA Ecosphere: A panoply of constituencies



Of course, different constituencies have different concerns

Everyone (*should*) worry about

- Accuracy of media reports
- Increasing over-interpretations & misinterpretations

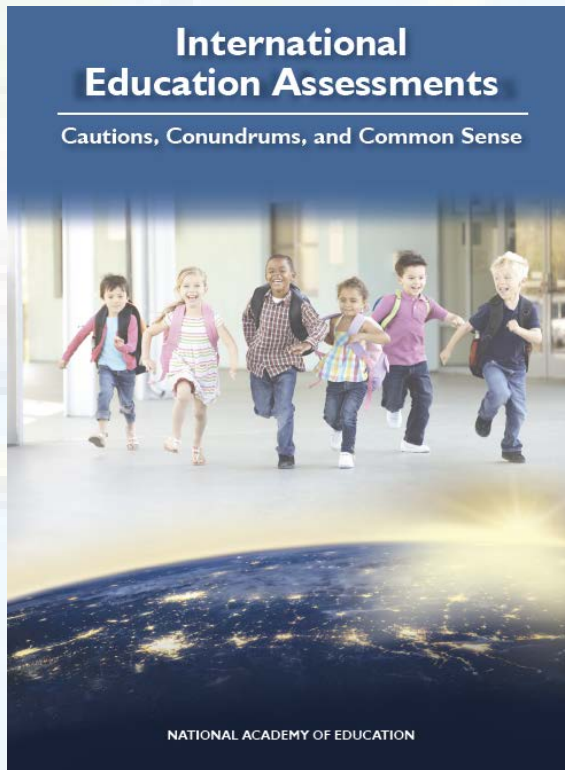
Policymakers worry about

- Increasing economic competition & insecurity
- Developing human capital
- The relevance of results to domestic education policy

Producers worry about

- Capacity to add countries & jurisdictions
- Impact of increasing participation for data quality
- Challenges of increasing heterogeneity in languages, cultures, and economic development
- Trustworthiness of cultural & linguistic adaptations

The report addresses *all* ILSA constituents & concerns ...
But today we will focus on three topics



1. Six Purposes of ILSAs
2. Selected Examples of Reporting & Interpretation
3. Policy Uses & Limitations
4. Design Issues
5. Analysis
6. Six Recommendations
(the report includes many more)

The Six Purposes of ILSAs

REALITY: The RQs we often *want* to answer \neq the RQs that we *can* answer (well)

	Purpose	How well can ILSAs achieve this purpose?
1	To describe and compare student achievement and contextual factors (e.g., policies, student characteristics) across nations	Excellent
2	To track changes over time in student achievement, contextual factors, and their mutual relationships, within and across nations	Excellent
3	To disturb complacency about the a nation's education system and to spur education reforms	★ Outstanding ★
4	To create de facto international benchmarking , by identifying top-performing nations and jurisdictions, or those making unusually large gains, and learning from their practices	With caveats
5	To evaluate the effectiveness of curricula, instructional strategies, and education policies	With extreme caution
6	To explore causal relationships between contextual factors (demographic, social, economic and educational variables) and student achievement	Dangerously difficult

How are ILSA results generally presented?

RANKINGS...and then, their purported implications

U.S. Students Slide In Global Ranking On Math, Reading, Science



Asia tops biggest global school rankings



The biggest ever global school rankings have been published, with Asian countries in the top five places and African countries at the bottom.

The UK is in 20th place, among higher achieving European countries, with the US in 28th.

Why Singapore's kids are so good at maths

The city-state regularly tops global league tables. What's the secret of its achievement?



Britain Turns to Chinese Textbooks to Improve Its Math Scores



And how do the top Asian countries cover ILSAs?: 2015 PISA

Japan

国際学力テストの日本の順位と平均得点

※丸数字は順位。数字は平均得点



Chinese Taipei

2015 PISA結果公佈 閱讀素養倒退7年
 The results of PISA 2015 are released. The reading achievement *[actually ranking]* regressed to 7 years ago *[actually, the score was higher]*

相信嗎？台灣PISA閱讀排名大倒退是因民族意識弱
 Do you believe it? The falling of Taiwan in PISA reading is due to a lack of patriotism

[Thanks to Xue Jiang and Shinji Katsumoto, Teachers College]

Yet ILSA results have led to important policy responses

Ireland's 2009 PISA Shock

Precipitous declines in reading and math scores spurred a major investigation that yielded a multi-factor diagnosis and political support for a “Literacy and Numeracy Strategy” that led to much-needed (but long-delayed) changes in

- Teacher education
- Curriculum and assessment
- School inspections and evaluations

England's 2012 PISA Shock

Strong showing of East Asian countries in math provided impetus for the mass adoption of their mastery textbooks for half of England's schools (2016)—More on that later.....

Key Message 1

Improve ILSA Dissemination

- **ILSA producers—sponsors and contractors alike—must work to counter the over-emphasis on league tables to the neglect of better ways of conveying**
 - Results in light of relevant contextual factors
 - Informative patterns of relationships among outcomes and contextual factors
- **Producers should devote greater resources to preparing reports—and reporters—providing more guidance on what can, and cannot, be inferred from results (media kits)**
- **New strategies are needed to convey on a regular basis the results and implications of high quality secondary analyses**

Key Message 2

Establish a National Advisory Board

- Build on the history of the **Board of International Comparative Studies in Education (BICSE)**
- Some illustrative potential board roles:
 - Advise NCES on ILSA goals and strategy; serve as a friendly critic with the goal of improvement
 - Offer independent cooperative assessments of ILSA results and their import for the US
 - Issue periodic reports on implications of ILSA-related research
- Agnostic about who might sponsor such a board—NCES, NAEd, NAS Committee on National Statistics—but the consensus was that this was worth doing

Key Message 3

Experiment with longitudinal follow-ups

- **ILSAs assess status; learning is about change**
- The report presents illustrative **longitudinal studies from other fields** (health & human development) that serve as promising examples for education
- **Conduct the necessary pilot studies** to evaluate whether and, if so, how ILSAs could be redesigned to include longitudinal components
- **In addition to encouraging countries to follow their own ILSA samples** (as the US, Germany, and Canada have done), for maximum pay-off **consider cross-national coordination of longitudinal follow-ups**

Key Message 4

Seize the opportunity of digitally based assessments (DBAs)

The transition to DBAs will have long-term consequences for ILSA design, utility and impact:

- **Improved cognitive instruments**
 - Better coverage and accuracy
 - Validation tools (e.g. effort tracking)
- **Improved background questionnaires**, with better coverage through
 - Adaptation
 - Rotated designs
- **Improved field operations**
 - Efficiency
 - Quality monitoring

Key Message 5

Link to extant data when distributing data files for analysis

- Even with improved background questionnaire data, ILSAs would be more **useful if files for secondary analysis included data from other sources**—especially measures of SES—such as Census data or administrative data collected by districts
- This is likely easiest in economically well-developed countries, where reliable measures related to school district SES can be derived using sophisticated approaches such as the U.S. Census’s small area income and poverty estimates
- **Links to extant data would help standardize secondary analyses**, improving their value and may even speed up the timetable for completion

Key Message 6 (from Singer & Braun, 2018, *Science*)

Use ILSA data to stimulate randomized field trials (RFTs)

- **Causal inference: The Holy Grail.** Whether any statistical analysis can yield inferences approximating causal statements was a matter of great debate during the workshops. Because we were not a consensus panel, we did not (attempt to) reach consensus.
- Although ILSA sponsors have conducted RFTs on their *methods*, we know of no large scale example to date of an RFT stimulated by ILSA results
- Consider what could have been learned were Britain to have conducted RFTs--perhaps with other English speaking countries—before rolling out mastery maths textbooks in so many of their schools



Despite these many cautions, **We believe the future of ILSA data is bright**

It is clear that ILSAs have:

- Helped us understand patterns of important relationships in cross-cultural contexts
- Led to constructive policy initiatives
- Spurred countries to action

Advances in design, development, delivery, analysis and reporting are coming, and will demand more research that will lead to:

- Improvements in domain coverage and accuracy
- Opportunities to better leverage the global investment in ILSAs
- More effective communications strategies

Everyone involved in this NAEd project hopes this report will spur efforts to strengthen future ILSAs (not damn them)

It is our collective responsibility... Join us!

Many thanks to NCES for funding this project.

