

# The Long Road From Evidence to Policy: Prospects for International Large-scale Assessment

Henry Braun

Lynch School of Education, Boston College

ILSA Conference

Educational Testing Service

March 16-18 2011

# The Landscape

- ILSA is a growth industry (See compilations by Hanushek and Woessmann)
  - Numbers of participating countries
  - Media attention
  - Impact
- Trajectory at an inflection point
  - Can growth be sustained?
  - Can relevance be enhanced?
- Opportune time to take stock of what has been accomplished – and what are options for the future
- Presentations have helped us in both respects

# Transparency (1)

- Transparency usually refers to “shining a light” on governmental procedures
- Ritzen focuses on illuminating national educational outcomes through international comparisons
- He suggests different mechanisms by which ILSAs can serve as change agents
- Much-needed key to unlock “hermetically sealed” educational systems

## Transparency (2)

- Is the impact of this kind of transparency always positive (or at least neutral)?
- Is transparency with respect to outcomes sufficient?
- How have different ILSAs evolved to
  - Address criticisms
  - Enhance their contributions to constructive change?

There are many faces to transparency ....







# Education Policy

The role of education policy is to design and manage the education system (including resource allocation) in order to achieve the goal of preparing all children to lead productive, satisfying lives that contribute to the common good.

Through what channels can/do policy makers learn to make improvements to the system?



# Questions to be Addressed

- What are the essential skills, dispositions and habits of mind that are required?
- What is possible with regard to growth and attainment of skills?
- What are the common characteristics of successful systems?

# ILSA: Theory of Action

Generating and disseminating credible evidence on the relative performance of different jurisdictions on meaningful constructs should:

- Spur reflection and review,
- Provide the impetus to consider modifications to policy and practice,
- Support the determination to allocate sufficient resources to achieve sustainable improvement.

# Theory of Action: Essential Conditions for Success

1. Credibility of assessment outcomes and cross-national comparisons
2. Recognition of important links between these outcomes and national goals
3. Governmental resolve and capacity to respond
4. Approximately accurate interpretations of results/comparisons
5. Formulation of appropriate policies
6. Sustained but flexible focus on those policies (reform initiatives)

# Credibility

- Broad consensus on operational definitions of target constructs
- Increasingly sophisticated test development procedures and psychometric analyses to ensure
  - Adequate construct representation
  - Comparability across countries and through time
- Other important factors include:
  - Translation
  - Motivation
  - Sample selection and participation

# Relevance

- Hanushek & Woessmann argue that core cognitive skills are key components of human capital (HC)
- Investment in HC brings both private and public benefits
- Country-level variation in HC related to variation in growth/development
- Importance of direct measures of skills vs. statistics on educational attainment
- Acknowledgement that returns to skills varies by country (level of development, etc.)

## Credibility and Relevance: Next Steps

- Depth: Improve representation of core constructs (complex skills) through appropriate use of technology, as well as other means
- Breadth: Increase range of constructs
  - ICT
  - Civic Engagement & Citizenship
  - Non-cognitive skills

# Civic Engagement

- Torney-Purta & Amadeo argue for looking beyond economic considerations to measuring dispositions related to civil society and participatory democracy
- Provide a useful review of past assessments and understandings generated through secondary analyses
- Balance between cross-national and sub-national comparisons
- Important insights generated through identification and description of key subpopulations

# Non-cognitive skills

- Levin argues for assessment of non-cognitive skills (to complement cognitive skills) because of importance (both theoretical and empirical) to success in school and work
  - Research of Heckman and others
  - Employer surveys
- Evidence that schooling influences the development of these skills
- Under-representation of these constructs can skew policy choices



# Possibilities for the Near-term Future

- Proof-of-concept: Inventory of “Student Approaches to Learning” in PISA 2000 (Marsh et al.)
- Substantial current work on improved instrumentation
- Potential of technology-based assessment
- Low-stakes of ILSAs may be well-suited to such measurement

# Technology

- Beller offers a useful comprehensive review of technology initiatives at the national and international levels
- Multiple pilots underway or planned
- Goals for introduction of technology
  - Facilitate assessment of new measures
  - Strengthen linkages to the world outside of schools
  - Enhance accuracy and access
  - Improve cost-effectiveness through new methodologies for development/ administration/ scoring/ reporting
- Formidable challenges in conducting technology-based administrations internationally

# Lessons from PIAAC

- PIAAC represents a major step forward in the use of technology
- Many lessons (some painful) learned on conducting an ILSA on a new technology platform
- Infrastructure can be leveraged for other purposes
- Hoped-for success of PIAAC should provide impetus for next generation of technology-based ILSAs
- Continue tradition of linking segments of the score scale to interpretable functionalities

# Governmental Resolve and Activity

- Varies by country and period due to interaction of multiple factors -- including “ambition”, political context, competing priorities
- Increasing prominence of ILSA results makes it harder for countries to dismiss or disengage
- Good intentions must be matched by capacity and sustainability
- ILSA sponsors can provide support and encouragement to within-country champions

# Interpretation and Formulation (1)

- Participating countries form a “natural experiment” but causal attributions are problematic because of formidable technical difficulties
  - Limitations of cross-sectional studies (Klieme)
  - Confounding of schooling differences with differences in other relevant conditions
  - High rankings or rapid progress likely due to a confluence of multiple factors
- Drawing general conclusions on “winning strategies” often not very convincing (cf: McKinsey reports)
- Using Canada as an example, Hargreaves argues that social and community conditions may be as or more important than specific schooling systems

## Interpretation and Formulation (2)

- Klieme's presentation provides a thoughtful analysis:
  - The difficulties in making inferences from ILSAs that are directly relevant to policy makers
  - Impossibility of carrying out credible value-added analyses
  - Potentially productive dynamic between ILSAs and Education Effectiveness Research
  - Examples from Germany on how to leverage ILSAs through addition of a longitudinal component
- Although initial ILSA reports are rich and encyclopedic, there is great value in focused secondary analyses
- Need more attention to sub-national comparisons

# Sustained Focus

- Essential given time lag in implementation of change in education systems
- Depends on broad political consensus on a long-term plan
- Periodic assessments assist in maintaining visibility
- Strategies should incorporate intermediate milestones whose attainment can maintain interest and support

# A Bright Future ... With Clouds on the Horizon?

As the number of participants grows, the burden on program staff multiplies ...

Can they continue to achieve broad consensus, preserve quality and meet tight time lines?

Is there a planning effort to explore alternative strategies for the future?

Can ILSAs become victims of their own success?



# The Sad Tale of the Irish Elk ...



# How Can the Positive Impact of ILSAs be Increased?

1. Provide more useful information
2. Enhance the value of that information
3. Extend the reach of ILSAs to more jurisdictions

Forward-looking strategies should encompass all three paths

# Information: More Useful, Enhanced Value

- Extend the range and depth of constructs assessed
  - Will often involve the use of technology
  - Greater use of special studies
- Provide more interpretable descriptions of different levels of performance
  - Behavioral anchoring
  - Modal student Models
- Use IALS as the “anchor” for time-linked complementary studies
  - TIMSS teacher video study
  - OECD leadership policy study (2006)
  - OECD teacher study (2008)
  - Longitudinal studies (Germany)
- Encourage links to national assessments
  - TIMSS-NAEP links

# Extending the Reach

- Partnerships with regional consortia
  - Sharing resources and expertise
  - Structuring intermediate levels of participation
- Building capacity in developing world
  - Technical expertise
  - Making policy use of data
- Media strategy
  - Continue high-profile events in different countries and various venues
  - Educate the press with respect to use of ILSA data

# Looking Ahead

ILSAs represent a (the?) major international educational commitment for many countries. It is imperative that the leaders of these initiatives strive to adapt the assessments to the evolving needs of the different countries while preserving sufficiently strong links to maintain trends.

Negotiating these sometimes conflicting desiderata will call on the best skills of both the measurement community and associated technical specialties. Good luck to us all!