Large-scale Learning Assessments

Where are we now?
What have we learned?
Where do we go from here?

Dr John Cresswell
NEQMAP Capacity Building Workshop
Bangkok, September, 2014
Overview

- Where are we now
  - International and regional assessments
  - National assessments and examinations

- What have we learned
  - How does assessment contribute to reform
  - Impact of studies globally, locally

- Where do we go from here
  - International, regional, national?
  - Setting up an assessment
This man wants to be rich so he is counting his money
Change happens - Korea

• In Korea, in the 1950s, the literacy rate was 22% and less than 20% of children attended secondary school.

• Education reforms undertaken by Korea led to them having one of the most advanced education systems in the world.
Change happens - Finland

- Until the late 1960s, in Finland most children left public school after six years. Some children went to private schools, academic grammar schools or folk schools. At that time only the privileged and fortunate were exposed to quality education.

- A series of reforms was undertaken over many years and in 2000 Finland achieved at the highest level in PISA.
Reasons for Collecting Assessment Data

- **Certification**
  - Examinations systems
  - Accreditation systems

- **Evaluation**
  - Teacher assessments to inform instruction
  - National, regional and international surveys
  - Effectiveness of initiatives, interventions and practices
Regional and International Assessments

- In 1964, the International Association for the Evaluation of Educational Achievement (IEA) conducted the first internationally comparative study in mathematics; 12 countries participated.

- By 2009, about 70 per cent of the countries in the world participated in some form of regional or international assessment program.

- Since the 1980s, the number of national assessment programs has also increased.
International Assessment Programs

- The International Association for the Evaluation of Educational Achievement (IEA)
  - PIRLS (reading Grade 4)
  - TIMSS (maths, science Grades 4 and 8)
  - ICCS (Civics)
  - ICILS (ICT literacy)
International Assessment Programs

- Organisation for Economic Cooperation and Development (OECD)
  - PISA (maths, reading, science, financial literacy, problem solving); 15-year-olds; since 2000; 72 participating systems in 2015.
- US AID/RTI – Early Grade Reading Assessment (EGRA) and Early Grade Mathematics Assessment (EGMA) – implemented in over 70 countries in 100 languages.
Regional Assessment Programs

- Southern and Eastern African Consortium for Monitoring Educational Quality (SACMEQ)
  - 15 countries; since 1995.

- Latin American Laboratory for the Evaluation of Quality in Education (LLECE)
Regional Assessment Programs

- The Conference of Education Ministers of Countries Using French as the Language of Communication (CONFEMEN)
  - Program for the Analysis of Educational Systems of CONFEMEN (PASEC); first assessment 1991; 24 participating countries in a series of assessments.
Regional Assessment Programs

- The Annual Status of Education Report (ASER): implemented annually in India for the last ten years, 600,000 out-of-school children.
- Uwezo: an ASER based assessment in Africa.
This man wants to be rich so he is counting his money again.
What do countries learn by participating?

- Gauge of educational health
  - Calls to action when results surprise
    - PISA: eg Germany, Japan
  - Setting of national expectations that are consistent with performance throughout the world
What do countries learn by participating?

- Influence on curriculum orientation and content
  - TIMSS (the role of statistics, probability and data in the mathematics curriculum)
  - PISA and its literacy orientation
  - Non-continuous text as an element of reading

- Capacity development
  - Assessment task development
  - Research design, sampling, scaling, analysis...
What have we learned?

What do countries learn by participating?

- Trend information
  - Is our performance increasing or decreasing?
  - Have gender differences narrowed?
  - Has participation of girls increased?

- Policy development
  - Inform options for policy
  - Identify correlates of outcomes (e.g., teacher qualifications, resource access)
This man wants to be rich so he is counting his money again.
Survey of Impact of Assessments

What have we learned?
Assessment Program Types

- National assessments were most frequently mentioned, largely in South America and Asia
- Regional assessment studies concentrated in Africa – South of Sahara (SACMEQ and PASEC)
- Sub-national assessment examples only found in larger countries of South America (e.g. Chile, Argentina, Brazil) and South Asia (e.g. India)
Stated Goals of Assessment

% of total mentions

- Quality: 34%
- Equity: 20%
- Accountability: 19%
- Leverage: 12%
- Other: 10%
- Unknown/NA: 5%

‘Other’ goals:
- Inform future assessments and build technical capacity
- Enable broad international comparisons
- Provide inputs to be used at local- or school-level
- Evaluate the effects of specific policies

What have we learned?
What have we learned?

Types of policy impacted

- **Resource Allocation**: 91
- **Teaching and Learning Practices**: 35
- **Curriculum Standards**: 23
- **OTHER**: 22
- **Assessment Policy**: 19
- **Performance Standards**: 18
- **Accountability Policy**: 10
- **Community/Parent Engagement Policy**: 7
- **Unknown/Not Applicable**: 6

- In 65% of studies involving international assessment programs, they were found to impact **curriculum standards**
- ‘Other’ policies impacted included well-being funding and interventions, performance-based pay and incentives, and expansion of early childhood education
**Resource Allocation Decisions**

<table>
<thead>
<tr>
<th>Category</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Instructional Materials</td>
<td>15</td>
</tr>
<tr>
<td>In-Service Professional Development</td>
<td>15</td>
</tr>
<tr>
<td>Funding Formula</td>
<td>12</td>
</tr>
<tr>
<td>Teacher Preparation</td>
<td>9</td>
</tr>
<tr>
<td>Teacher Recruitment and Retention</td>
<td>7</td>
</tr>
<tr>
<td>Facilities</td>
<td>7</td>
</tr>
<tr>
<td>Decision Making Authority</td>
<td>4</td>
</tr>
<tr>
<td>Instructional Time/School Hours</td>
<td>4</td>
</tr>
<tr>
<td>Equipment</td>
<td>4</td>
</tr>
<tr>
<td>Class Size/Ratios</td>
<td>2</td>
</tr>
<tr>
<td>School Supplies</td>
<td>2</td>
</tr>
<tr>
<td>School Feeding/Meals</td>
<td>1</td>
</tr>
</tbody>
</table>

- The most common resource allocation decisions were ones intended to improve the quality of the teacher and teaching materials used in schools.
What are some issues with the assessments

- Under-utilisation of the data
  - Both national and international
  - Although it can be a conversation starter
- Unintended outcomes
  - Curriculum narrowing
  - Possibility of teaching the test
- Interpretation without a full understanding of context
- Setting unobtainable goals.
What are some issues with the assessments

- The political life-cycle is much shorter than the education reform life-cycle.
- Lack of relevant understandable dissemination of results. Especially to the public where press reports will focus on the negative aspects on the results.
- Erroneous and/or inappropriate use of the data
What are some issues with the assessments

• The cost of conducting a large-scale assessment is large – lack of funds can affect the ability to collect and analyse meaningful data for policymakers.
• Although it can also be said that the cost of learning assessments is relatively small when compared to the cost of policy implementation and the cost of making uninformed, incorrect decisions.
• Funding and political insecurity might prevent continuity and regularity.
This man wants to be rich so he is counting his money again.
In summary

If learning assessments are implemented, they should involve the systematic and strategic collection of data on education outcomes and factors related to those outcomes and be used to inform policy and practice aimed at improving progress for all learners.

Focussed on where learners are at, not where we want them to be.

We need data on outcomes – in addition to inputs.

regularity & sustainability

Where do we go from here?
In summary

If learning assessments are implemented they should involve the systematic and strategic collection of data on education outcomes and factors related to those outcomes and be used to inform policy and practice aimed at improving educational progress for all learners and growth.

Instrumentation and methodology better oriented to informing practice and policy.

Where do we go from here?
This man wants to be rich so he is counting his money again
Core Goal of the Learning Metrics Task Force

The overarching goal of the Learning Metrics Task Force is to catalyse a shift in the global conversation on education from a focus on access to access plus learning.

Assessment is a necessary part of the process to find out if students are learning.
This man wants to be rich so he is counting his money again.
Assessment is just one part of the process of educational improvement.

Simply doing more assessment will not, by itself, improve educational outcomes of students.

Just like the man assessing his wealth by counting the money over and over again will not make him rich - he needs to undertake some other activities.
Reform cycle
Education reform is not a straight line activity

Reform discussion

Student assessment and reporting

Policy dialogue and decisions

Policy implementation

Teaching and learning

Where do we go from here?
Elements of an effective assessment program

- To ensure that an assessment program contributes to policy reform and improved student outcomes, it should have a number of interlinked elements.
Elements of an assessment program

- Reporting and dissemination
- Linkage to policy response
- Project team and infrastructure
- Policy issues for the assessment
- Technical standards
- Assessment framework
- Assessment items
- Linguistic quality control
- Data analysis
- Scaling methodology
- Standardised field operations
- Sample design
- Test design

Where do we go from here?
Elements of an assessment program

- Reporting and dissemination
- Linkage to policy response
- Project team and infrastructure
- Policy issues for the assessment
- Technical standards
- Assessment framework
- Assessment items
- Linguistic quality control
- Sample design
- Test design
- Standardised field operations
- Scaling methodology
- Data analysis

Where do we go from here?
Project team and infrastructure

- Staff are required to cover the roles of project management, policy liaison, assessment and item development, translation coordination, field operations, project administration, data management, sampling, analysis, reporting and communications.
- In addition to staffing, adequately resources and a core physical infrastructure is needed including computing and software support.
Policy issues for the assessment

- The goal of an assessment is to estimate student outcomes in a way that informs educational policy and practice.
- Policy makers and practitioners are more likely to use data from an assessment to inform educational policy if the assessment prioritises their concerns.
Technical standards of Implementation

- Technical standards reinforce the consistency, precision and generalisability of the data. Typically, standards are created for operational areas such as school and student response rates, field trial sample size, psychometric quality of the instrumentation, security of items, and test administration procedures.
Assessment framework

- An assessment framework is an explicit statement and discussion about what an assessment intends to measure and it lays out the principles upon which an assessment is built. It guides test development.
High quality assessment items

- Assessment items are the heart of an assessment program and careful attention must be paid to their development and quality. Item development is both a science and an art.
- If possible, field trial items with the target population.
Linguistic quality control

- Where assessments are administered in different languages, the different versions should be equivalent.
- Double independent translation is recommended.
Test design

- The test design will usually specify the format of materials; the number of test forms; how the materials are rotated and linked between booklets; and the balance of contextual and cognitive questions.
- Is the test computer-based, paper-based, or delivered on a tablet or a mobile phone.
Sample design

- This requires careful specification of target population(s) and techniques for accessing random samples from those populations with known sampling probabilities.
- Is it sample or census? Is it grade-based or age-based?
- Does your organisation have the capacity to do the weighting?
Standardised field operations

- Typically, field operations manuals are produced which specify the length of the testing sessions, the instructions for the students, and the physical aspects of the testing environment.
- Ideally, the test administrators will be independent and therefore impartial with regards to student outcomes.
Scaling methodology

- Item-response theory is currently recognised as the preferred methodology for evaluating and ensuring that the assessment instruments are of appropriate technical quality.
- Scales are constructed from the assessment data using item-response theory.
Data analysis

- Using item-response theory for scaling, the calculation of plausible values is the standard for achieving unbiased population estimates.
- With complex surveys, the use of replication methods is required to accurately determine sample variance, and consequently significant differences between groups – does your staff have this capacity?
Reporting and dissemination

It is important to determine an appropriate strategy for disseminating the results to different audiences. Teachers and curriculum developers will have different foci, as will parents, policy analysts and economists.
Linkage to policy response

- Relating the assessment back to the policy issues determined at the outset of the development process adds value to the program.
- A well-designed assessment program includes a follow-through to inform and influence a policy response. This activity goes beyond dissemination and can include an analysis of policy options.
Burning Questions

GETTING STARTED

• How do we obtain funding for assessments?
• Would UNESCO implement a learning assessment?
• How do I introduce large-scale assessment into my country?
• Do people in this region have a different view of assessment than other parts of the world?
• I want to know more about how to conduct a large-scale assessment.
• I want to learn about experiences in other countries.
• What is the right assessment for my country – is the context right?
Burning Questions

THE ASSESSMENT

- How do you measure every aspect of learning with a paper and pencil test?
- How do you cope with such a wide range of abilities in a national assessment?
- Is assessment authentic in a large-scale assessment?
- How do you assess different types of students?
- Can the integration of technology enhance the assessment experience?
- Who monitors the assessment?
- How do you develop standards for different levels of education?
- How do you develop standards in education?

Where do we go from here?
Burning Questions

USING THE ASSESSMENT

• How do we make people aware of the findings of assessments?
• How do countries use assessment to bring about policy change?
• How do you use large-scale assessment to improve teaching?
• How do you use large-scale assessment to influence classroom teaching?

LAST BURNING QUESTION

• What happened to the man who was counting his money?
Based on the assessment that his money was not increasing, this man, who wanted to be rich, undertook a program of careful savings and investment.
Now he is so rich he needs help to count his money!
Thank you

john.cresswell@acer.edu.au
Thank you

john.cresswell@acer.edu.au