Session 7: Introduction to international large-scale assessments

13:30 till 15:00 Day 2

Presented by Jim Tognolini
Introduction to international large-scale assessments

During this session we will

• introduce international assessment systems (e.g. PISA, PISA_D, TIMSS, PIRLS, EGRA and ASER);

• describe the basic characteristics of such systems;

• show how they align to modern assessment theory: and,

• conduct a workshop session where groups review and reflect upon each of the international programs.
Presentation on international large-scale assessments

This session will comprise a number of components.

1. Participants will be located in one of 5 groups (PISA, including PISA D; TIMSS; PIRLS; EGRA; and, ASER).

2. Participants will be given my responses to some basic questions on the particular assessment to review and consider in their group.

3. Participants will then respond to the questions in the “Test Exercise”. These questions focus on clarification; seeking further information; looking at the advantages and disadvantages of the test for Bhutan; and, making a judgement (with a supporting case) about the suitability of the test for Bhutan.

4. Each group will then present to the total group in a plenary session.
The international programs

• The Program for International Student Assessment (PISA)
• The Program for International Student Assessment Development (PISA D).
• Trends in International Mathematics and Science Study (TIMSS)
• Trends in International Mathematics and Science Study (TIMSS) Numeracy (not covered in this analysis)
• Progress in International Reading Literacy Study (PIRLS)
• Progress in International Reading Literacy Study (PIRLS) Literacy (not covered in this analysis)
• ePIRLS (not covered)
• Early Grade Reading Assessment (EGRA)
• ASER - The Annual Status of Education Report
The international programs
PISA

1. What does PISA stand for?
The Program for International Student Assessment (PISA)

2. What is assessed by the PISA Tests?
PISA assesses literacy in the context of subjects like reading, mathematics and science. The tests are not directly linked to traditional school curricula. They do assess students’ ability to apply knowledge and skills in key subject areas and to analyse, reason and communicate effectively as they examine, interpret and solve problem in a real life context. They do enable comparisons over time.

More recently, PISA has also tested **Collaborative Problem Solving** and **Financial Literacy** for those countries that wished to participate.

3. Which countries/economies participate in PISA?
There were 65 economies participating in 2012; and, 71 are signed up to participate in 2015.
4. What sort of tests are used?
So far pencil-and-paper tests have been used (although this changed for some countries in 2009 and 2012 and in 2015 the default mode of testing was computer). The tests are made up of both multiple-choice questions and extended response questions. The questions are usually linked to texts and sometimes includes pictures, graphs or tables setting out real-life situations. Each student takes a two-hour test, where the actual tests are different for every student.

5. Who writes the test questions?
Contractors employed by OECD, although all participating countries can contribute items and all will have the chance to review the items.

6. Are the PISA tests trialed before the final administration?
Yes. The items are trialed in all participating countries. If any test questions prove to have been too easy or too hard in certain countries/economies, they are dropped from the real test in all countries and economies.
7. **Are context questionnaires used?**
   Yes. Students answer a background questionnaire, providing information about themselves, their attitudes to learning and their homes. In addition, school principals are given a 20-minute questionnaire about their schools.

8. **How is the questionnaire data used?**
   It is used to help countries explore the relationship between how students perform in PISA and factors such as migration, gender and students’ socio-economic background, as well as students’ attitudes about school and their approaches to learning.

9. **Who marks the tests?**
   They are marked by the participant countries using centralised prepared marking keys and then randomly checked by senior global markers.
What do the test scores mean?
PISA scores enable the distribution of scores for the students from a country and for each subject to be located on a developmental continuum (scale) using a measurement model (Rasch Model). This developmental continuum shows the performance standards which are represented by levels. Where Level 1 indicates that students have achieved the most basic skills and higher levels show increasing levels of cognition.

What is the metric of the PISA Scale?
In each test subject, the score for each participating country is the average of all student scores in that country. The average score among OECD countries is 500 points and the standard deviation is 100 points. About two-thirds of students across OECD countries score between 400 and 600 points.
The international programs
PISA

12. Who is assessed by PISA?
Schools in each country are randomly selected for participation in PISA. At these schools, the test is given to students who are between age 15 years 3 months and age 16 years 2 months at the time of the test, rather than to students in a specific year of school. The average age across the sample is approximately 15 years.

13. Who pays for PISA?
PISA is financed directly by the participants' government authorities, typically education ministries.

14. Are the results provided made available by school and/or students?
No. The way the data are collected data collected (using sampling) really only allows countries and economies to compare themselves to pre-determined standards (standards referencing) AND to each other (norm-referencing). Systems can enter more students and make other comparisons within their own country.
The international programs
PISA Exercise

1. Are there any clarifying questions from the questions and answers provided? If yes, what are they?

2. Are there any further questions regarding PISA?

3. What are the advantages and disadvantages are of Bhutan joining the PISA Program?

4. Rate your response to the following statement and give your reasons why?

<table>
<thead>
<tr>
<th>Statement</th>
<th>Strongly Agree</th>
<th>Agree</th>
<th>Undecided</th>
<th>Disagree</th>
<th>Strongly Disagree</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall I consider that Bhutan should participate in PISA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

5. Is there anything else we should know about PISA?
1. What does PISA D stand for?
   PISA for Development

2. What is PISA for Development?
The OECD’s PISA for Development initiative involves developing enhanced PISA assessment instruments in reading, mathematics and science literacy; methods; and, analyses to make PISA more relevant to developing countries but that produce comparable scores as the main PISA assessment.

3. Which countries are participating in the initial trial?
   Cambodia, Ecuador, Guatemala, Honduras, Senegal, Sri Lanka and Zambia.

4. Are there technical issues with carrying out this initiative?
   Yes. There are technical papers which suggest that the equating could be problematic.
The international programs
PISA D Exercise

1. Are there any clarifying questions from the questions and answers provided? If yes, what are they?

2. Are there any further questions regarding PISA D?

3. What are the advantages and disadvantages of Bhutan joining the PISA D Program?

4. Rate your **response to the following statement and give your reasons why?**

<table>
<thead>
<tr>
<th>Statement</th>
<th>Strongly Agree</th>
<th>Agree</th>
<th>Undecided</th>
<th>Disagree</th>
<th>Strongly Disagree</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall I consider that Bhutan should participate in PISA D</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

5. Is there anything else we should know about PISA D?
The international programs TIMSS

1. What does TIMSS stand for?
The Trends in International Mathematics and Science Study (TIMSS).

2. What is assessed by the TIMSS Tests?
TIMSS assesses the mathematics and science knowledge of students around the world in the fourth and eighth grades. The first TIMSS assessment conducted in TIMSS in 1995 and has continued every four years—1999, 2003, 2007, 2011 and 2015. TIMSS is conducted by the International Association for the Evaluation of Educational Achievement (IEA).

The TIMSS frameworks and specifications are developed in a collaborative process and include the interests of a wide range of countries.

3. Which countries/economies participate in TIMSS?
More than 60 countries and systems participated in TIMSS in 2015. A random sample of students is used in each country.
The international programs
TIMSS

4. What sort of tests are used?
So far pencil-and-paper tests have been used. The tests are made up of both multiple-choice questions (about 60% in Grades 4 and 8 in mathematics) and extended response questions (about 40% in Grades 4 and 8 in mathematics). In science about 55% of the items are multiple-choice and 45% extended response in both Grades 4 and 8. Each student takes approximately 1 and a quarter hour test followed by a 30-minutes context questionnaire.

5. Who writes the test questions?
Contractors, although all participating countries can contribute items and all will have the chance to review the items.

6. Are the TIMSS tests trialed before the final administration?
Yes. The items are trialed in all participating countries. If any test questions prove to have been too easy or too hard in certain countries/economies, or function poorly, they are dropped from the real test in all countries and economies.
7. Are context questionnaires used?
Yes. Background questionnaires are administered to students, their teachers, and their school principals. In 2015, the fourth grade TIMSS assessment included a home questionnaire for students’ parents and caregivers. This questionnaire is designed to collect information about students’ home backgrounds and early learning experiences. Curriculum questionnaires are also administered to experts to collect information about educational policies and the national contexts that shape the content and implementation of the mathematics and science curricula across countries.

8. How is the questionnaire data used?
It is used to help countries explore the relationship between how students perform in TIMSS and factors such as migration, gender and students’ socio-economic background, as well as students’ attitudes about school and their approaches to learning.

9. Who marks the tests?
They are marked by the participant countries using centralised prepared marking keys and then randomly checked by senior global markers.
10. What do the test scores mean?
TIMSS scores enable the distribution of scores for the students from a country and for each subject to be located on a developmental continuum (scale) using an IRT 2-parameter measurement model. This developmental continuum shows the performance standards which are represented by levels. Where Level 1 indicates that students have achieved the most basic skills and higher levels show increasing levels of cognition. The results from Grades 4 and 8 are placed on the same scales and results from administration to administration can be also directly compared.

11. What is the metric of the TIMSS Scale?
In each test subject, the score for each participating country is the average of all student scores in that country. The average score among participating countries in the first year of administration was 500 points and the standard deviation was 100 points.
12. **Who is assessed by TIMSS?**

Schools in each country are randomly selected for participation in TIMSS. At these schools, the test is given to students who Grades 4 and Grade 8 at the time of the test.

13. **Who pays for TIMSS?**

TIMSS is financed directly by the participants' government authorities, typically education ministries.

14. **Are the results provided made available by school and/or students?**

No. The way the data are collected really only allows countries and economies to compare themselves to pre-determined standards (standards referencing) AND to each other (norm-referencing). Systems can enter more students and make other comparisons within their own country.
1. Are there any clarifying questions from the questions and answers provided? If yes, what are they?

2. Are there any further questions regarding TIMSS?

3. What are the advantages and disadvantages are of Bhutan joining the TIMSS Program?

4. Rate your *response to the following statement and give your reasons why?*

<table>
<thead>
<tr>
<th>Statement</th>
<th>Strongly Agree</th>
<th>Agree</th>
<th>Undecided</th>
<th>Disagree</th>
<th>Strongly Disagree</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall I consider that Bhutan should participate in TIMSS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

5. Is there anything else we should know about TIMSS?
The international programs
PIRLS

1. **What does PIRLS stand for?**
The Progress in International Reading Literacy Study (PIRLS)

2. **What is assessed by the PIRLS Tests?**
PIRLS assesses reading literacy in the fourth grade (represented by being in their fourth year of schooling). Reading literacy is currently defined as "the ability to understand and use those written language forms required by society and/or valued by the individual. Readers can construct meaning from texts in a variety of forms. They read to learn, to participate in communities of readers in school and everyday life, and for enjoyment".

The first administration was in 2001 and occurs every 5 years.

The PIRLS frameworks and specifications are developed in a collaborative process and include the interests of a wide range of countries.

3. **Which countries/economies participate in PIRLS?**
About 40 countries and systems participated in PIRLS in 2011. A random sample of students is used in each country.
The international programs
PIRLS

4. What sort of tests are used?
   So far pencil-and-paper tests have been used. The tests are made up of both multiple-choice questions (about 45% in 2011) and extended response questions (about 55% in 2011). Each student takes approximately 1 and a quarter hour test followed by a 30-minutes context questionnaire.

5. Who writes the test questions?
   All participating countries can contribute items and all will have the chance to review the items.

6. Are the PIRLS tests trialed before the final administration?
   Yes. The items are trialed in all participating countries. If any test questions prove to have been too easy or too hard in certain countries/economies, or function poorly, they are dropped from the real test in all countries and economies.
7. Are context questionnaires used?
Yes. Background questionnaires are administered to students, their teachers, and their school principals. A home questionnaire for students’ parents and caregivers is given. This questionnaire collects information on "students’ early reading experiences, child-parent literacy interactions, parents’ reading habits and attitudes, home-school connections, and demographic and socioeconomic indicators."

8. How is the questionnaire data used?
It is used to help countries explore the relationship between how students perform in PIRLS and factors such as migration, gender and students’ socio-economic background, as well as students’ attitudes about school and their confidence in reading.

9. Who marks the tests?
They are marked by the participant countries using centralised prepared marking keys and then randomly checked by senior global markers.
10. What do the test scores mean?

PIRLS scores enable the distribution of scores for the students from a country and for each subject to be located on a developmental continuum (scale) using an IRT 2-parameter measurement model. This developmental continuum shows the performance standards which are represented by levels. Where Level 1 indicates that students have achieved the most basic skills and higher levels show increasing levels of cognition. The results from administration to administration can be directly compared. This allows monitoring over time.

11. What is the metric of the PIRLS Scale?

In each test subject, the score for each participating country is the average of all student scores in that country. The average score among participating countries in the first year of administration was 500 points and the standard deviation was 100 points.
The international programs

PIRLS

12. Who is assessed by PIRLS?
TIMSS and PIRLS both employ the same two-stage random sample design, with a sample of schools drawn as a first stage and one or more intact classes of students selected from each of the sampled schools as a second stage. Each country is responsible for drawing their own sample based on the parameters provided by the IEA.

13. Who pays for PIRLS?
PIRLS is financed directly by the participants' government authorities, typically education ministries.

14. Are the results provided made available by school and/or students?
No. The way the data are collected (using sampling) really only allows countries and economies to compare themselves to pre-determined standards (standards referencing) AND to each other (norm-referencing). Systems can enter more students and make other comparisons within their own country.
The international programs
PIRLS Exercise

1. Are there any clarifying questions from the questions and answers provided? If yes, what are they?

2. Are there any further questions regarding PIRLS?

3. What are the advantages and disadvantages are of Bhutan joining the PIRLS Program?

4. Rate your *response to the following statement and give your reasons why*?

<table>
<thead>
<tr>
<th>Statement</th>
<th>Strongly Agree</th>
<th>Agree</th>
<th>Undecided</th>
<th>Disagree</th>
<th>Strongly Disagree</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall I consider that Bhutan should participate in PIRLS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

5. Is there anything else we should know about PIRLS?
The international programs
EGRA

1. **What does EGRA stand for?**
   Early Grade Reading Assessment (EGRA).

2. **What is assessed by the EGRA Assessments?**
   EGRA is an oral assessment that tests the most basic foundation skills for literacy acquisition in early grades, including pre-reading skills such as listening comprehension. The test includes timed, 1-minute assessments of letter naming, nonsense and familiar words, and paragraph reading. Additional (untimed) segments include comprehension, relationship to print, and dictation.

3. **Which countries/economies participate in EGRA?**
   More than 40 countries and systems participated in EGRA and it is carried out in a host of languages.
The international programs
EGRA

4. What sort of tests are used?
EGRA is an individually administered oral assessment of foundation literacy skills requiring about 15 minutes per child. It has been designed as an inexpensive and simple diagnostic of individual student progress in reading.

5. Who writes the test questions?
The basic EGRA assessment is written by contractors. However, most countries usually adapt it including for translation into the language of instruction. As such the results are not generally able to be compared across countries. One of the reasons for changing the assessment is that the way these skills develop is different from one country to the next.

6. Are the EGRA tests trialed before the final administration?
Yes. In a limited way the basic assessment is trial tested.
The international programs

EGRA

7. Are context questionnaires used?
   As far as I can ascertain there are no context questionnaires used with EGRA.

8. How is the questionnaire data used?
   Not applicable

9. Who marks the tests?
   They are marked by the participant countries using centralised prepared marking keys.
10. What do the test scores mean?
EGRA is more a diagnostic assessment and as such most of the results provide the average number of students achieving some basic standard across year levels. These data are used to inform policy at a system level and also provide teachers with information that can be used to improve the foundation skills of students entering formal education. Research-based instructional programs have been developed in areas identified by the EGRA assessment, to help them improve learning.

11. What is the metric of the EGRA Scale?
Not applicable.
The international programs
EGRA

12. Who is assessed by EGRA?
For system level analysis, approximately 30 schools are chosen at random and then approximately 15 students are randomly chosen from within schools to sit the assessment. As with all of these studies the sample size is determined by the required confidence interval and the costs associated with collecting the data.

13. Who pays for EGRA?
EGRA is financed directly by the participants' government authorities, typically education ministries.

14. Are the results provided made available by school and/or students?
This can vary; however, when system wide information is required, the data that are collected are only used to provide summary data at a system level.
The international programs
EGRA Exercise

1. Are there any clarifying questions from the questions and answers provided? If yes, what are they?

2. Are there any further questions regarding EGRA?

3. What are the advantages and disadvantages are of Bhutan joining the EGRA Program?

4. Rate your **response to the following statement and give your reasons why?**

<table>
<thead>
<tr>
<th>Statement</th>
<th>Strongly Agree</th>
<th>Agree</th>
<th>Undecided</th>
<th>Disagree</th>
<th>Strongly Disagree</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall I consider that Bhutan should participate in EGRA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

5. Is there anything else we should know about EGRA?
The international programs
ASER

1. What does ASER stand for?

2. What is assessed by the ASER Tests?
ASER assesses basic reading and basic arithmetic for children aged between 5 and 16. The same test is administered to all children. It has also, on occasion, tested basic English and Everyday mathematics tasks.

3. Which countries/economies participate in ASER?
ASER is conducted in each state and rural province in India. It is also administered in Pakistan, Kenya, Tanzania, Uganda, Mali, Senegal, and Mexico.
**The international programs ASER**

4. **What sort of tests are used?**
   The tests are administered one-on-one in an oral format by volunteer administrators in the homes of the children chosen.

5. **Who writes the test questions?**
   Staff of the ASER Centre wrote the original assessments, and the same instrument is used every year.

6. **Are the ASER tests trialed before the final administration?**
   They were when the original tests were constructed.
The international programs

ASER

7. Are context questionnaires used?
Yes. Basic household information is collected every year. In recent years this has included household size, parental education, and some information on household assets. In addition, one government primary school in each sampled village is visited in order to collect information on school infrastructure, enrolment, attendance, teachers and fund flows.

8. How is the questionnaire data used?
It is used to help countries explore the relationship between how students perform in ASER and some background demographic factors such as gender and students’ socio-economic background.

9. Who marks the tests?
They are marked by the volunteer administrators.
10. **What do the test scores mean?**

ASER scores provide basic descriptive statistics at an item level. For example, “Half of all children in Standard V have not yet learned basic skills that they should have learned by Standard II”; and, “Close to half of all children will finish eight years of schooling but still not have learned basic skills in arithmetic.”

11. **What is the metric of the ASER Scale?**

Not applicable at this stage.
12. **Who is assessed by ASER?**
   In each rural (not urban) district, 30 villages are randomly sampled. In each village, 20 randomly selected households are surveyed. This process generates a total of 600 households per district, or about 300,000 households for the country as a whole. Approximately 600,000 children in the age group 3-16 who are resident in these households are surveyed.

13. **Who pays for ASER?**
   ASER is financed indirectly foundations, NGOs, charitable foundations, Aid Agencies, etc.

14. **Are the results provided made available by school and/or students?**
   No.
The international programs
ASER Exercise

1. Are there any clarifying questions from the questions and answers provided? If yes, what are they?

2. Are there any further questions regarding ASER?

3. What are the advantages and disadvantages are of Bhutan joining the ASER Program?

4. Rate your response to the following statement and give your reasons why?

<table>
<thead>
<tr>
<th>Statement</th>
<th>Strongly Agree</th>
<th>Agree</th>
<th>Undecided</th>
<th>Disagree</th>
<th>Strongly Disagree</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall I consider that Bhutan should participate in ASER</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

5. Is there anything else we should know about ASER?