Session 1: Introduction to Modern Assessment Theory: A basis for all assessments

10:30 till 12:30 Day 1

Presented by Jim Tognolini

25-27 APRIL, 2016 / PARO, BHUTAN
Introduction to Modern Assessment Theory: *A basis for all assessments*

During this session we will

- define the terms *assessment, testing, evaluation and measurement, summative assessment, formative assessment, school-based assessment, assessment-for and assessment-of learning*;
- introduce the concept of an *image in assessment*;
- consider ways that *marks are given meaning*;
- introduce the concept of a *developmental continuum*;
- articulate the purposes of *educational assessments*; and,
- discuss the *difference between item difficulty and cognition*.
<table>
<thead>
<tr>
<th>TERM</th>
<th>No. of Term for Definition</th>
<th>DEFINITION</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Formative Assessment</td>
<td></td>
<td>is the process of making a value judgement about the quality of the information collected.</td>
</tr>
<tr>
<td>2. School Based Assessment</td>
<td>2</td>
<td>refers to a system of teacher assessment of a student that covers all aspects of student development</td>
</tr>
<tr>
<td>4. Assessment</td>
<td></td>
<td>is one relatively formal way of collecting information about students’ abilities of completing tasks.</td>
</tr>
<tr>
<td>4. Summative Assessment</td>
<td></td>
<td>involves professional judgement based upon an image formed by the collection of information about student performance</td>
</tr>
<tr>
<td>5. Evaluation</td>
<td></td>
<td>is typically used to assess achievement at the end of a course.</td>
</tr>
<tr>
<td>6. Testing</td>
<td></td>
<td>is an assessment which is embedded in the teaching and learning process.</td>
</tr>
<tr>
<td>7. Measurement</td>
<td></td>
<td>is the process of seeking and interpreting evidence for use by learners and their teachers to decide where the learners are in their learning, where they need to go and how best to get there.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>is the process of ascribing a number which summarises how much of the property (construct) is present in the object.</td>
</tr>
</tbody>
</table>
Definition of assessment

Assessment involves *professional judgement* based upon an *image* formed by the collection of information about student performance.
The image

Teachers build images of what students know and can do based upon all the information that is collected from various assessment techniques.

The IMAGE is critical to teaching/learning and standard settings.
Assessment Types: Assessment for Learning and Assessment of Learning

- **Assessment for Learning** is the process of seeking and interpreting evidence for use by learners and their teachers to decide where the learners are in their learning, where they need to go and how best to get there. (Assessment Reform Group (2002))

- **Assessment of learning** is the process of evaluating the extent to which participants in education have developed their knowledge, understanding and abilities.
What meaning does a mark have?

What does a mark really mean?

If someone gets a mark of 88, what does it REALLY mean?
Interpreting Scores

- Marks are given meaning by referencing them to the performance of the cohort [Norm Referencing].
- Movement away from referencing marks to norms towards referencing them to criteria [Criterion Referencing] – some systems haven’t moved past here except in nomenclature.
- Currently there is a move towards referencing marks to pre-determined standards [Standards Referencing].
A new way of thinking about assessment

• Assessment is viewed as the process of monitoring a child’s progress along a *developmental continuum*.

• The focus shifts from comparing one individual with another, towards monitoring what students know and can do.

• It is a model based on *growth*.
The developmental continuum

A developmental continuum attempts to capture in words what it means to make progress or to improve in an area of learning or domain of knowledge.
Developmental sequence for Whole numbers

Orders, reads and writes numbers of any size

Counts, orders, reads and records numbers up to and including four digits

Counts, orders, reads and represents two- and three-digit numbers

Counts to 30 and orders, reads and represents numbers in the range 0 to 20.
Developmental sequence for problem solving

- Suggests, considers and selects appropriate alternatives when resolving problems
- Draws on past experience when solving problems
- Seeks help as needed when faced with simple problems
- Has little idea of how to go about solving problems
Developmental continuum for showing GROWTH in an area of learning.
Example developmental continuum for mathematics

\[ \int \sin(x-1)^n \, dx \]

How can you calculate the area of this shape?

What is the difference between the largest and second largest numbers that can be made from the following digits: 3, 7, 4 and 6?

76 - 29 =

17 + 29 =

2 + 2 =
The teachers and examiners task ...

... is to locate the student on the developmental continuum by constructing items, questions and tasks that give the students the opportunity to demonstrate what it is they know and can do in relation to the indicators which emerge from the standards. The further to the right the student is on the continuum, the more deep is the knowledge that they have in the area of the curriculum.
Difficulty level and cognition
Item difficulty

• For multiple choice questions the item difficulty is defined as the proportion of test takers who answer the item correctly. If everyone answers the item correctly the item difficulty is 1; if no-one answers it correctly the item difficulty is 0.

• It is usually referred to as the item’s p-value.

  ➢ Item would be referred to as HARD if the p-value is less than 0.30.
  ➢ Item would be referred to as MEDIUM if the p-value is between 0.30 and 0.70.
  ➢ Item would be referred to as EASY if the p-value is greater than 0.7.

• Generally a range of difficulty levels is required.

• The item difficulty is determined after the test is administered.
Item difficulty

• An examiner should write items with the intent that each item written will fall within one of the levels of difficulty.

• In order to determine difficulty levels, examiners should use their
  
  ➢ knowledge of the ability of the students;
  ➢ knowledge of the content that they are writing the items to assess; and,
  ➢ knowledge of types of misconceptions (common errors) that the students make on the types of items being written.
Varying the item difficulty

• You can vary the vocabulary.
  
  ➢ Instead of asking who is the *antagonist* in a play; you could ask who is the opponent? The latter question would be easier purely because of the vocabulary and this might be more appropriate.

• You can vary the familiarity of the content or context.

  ➢ Instead of asking for the *scientific names* of organisms; you could ask for the *common names* of organisms. The latter would be easier. Depending upon the learning objective this may or may not be appropriate

• You can vary the format.

  ➢ Instead of asking an addition question as follows 234 + 179; you could ask the following 234 + 179. The latter would be easier.
Varying the item difficulty

- You could add a diagram.
  - Instead of asking to find the circumference of a circle with a diameter of 5 cms; you could ask what is the circumference of the circle shown?
Quick Quiz (True or False or Don’t Know [Need More Information])

1. The p-value is used to determine the difficulty of an item.

2. An item answered correctly by a high proportion of test takers is more difficult than a different item answered correctly by a low proportion of test takers.

3. If examiners estimate that 75% of students will answer a question correctly, they should classify it as a hard item.

4. An examiner can make an item easier by using more common vocabulary in the item.

5. Each item on an assessment should have approximately the same p-value.

6. An item p-value can be expressed as a percentage.
Cognitive Level

• The cognitive level of an item refers to the complexity or demand of the thought processes required to correctly answer the item.

• Cognitive level relates to depth of knowledge assessed as an indication of progress along the developmental continuum.

• Generally, deeper cognitive level items would be expected to be harder but this needs to be checked using the p-value after the class has taken the test.
Writing to Cognitive Levels

• There are a variety of cognitive models and tools that can be used to categorize levels of learning.

• These models organise cognitive levels by the complexity of the thought processes required.

• These models use verbs to describe the complexity of the thought processes.

• The most widely used model or tool is Bloom’s Taxonomy; there are several versions of it.
Bloom’s Taxonomy

1950

Higher Order Thinking skills
- Evaluation
- Synthesis
- Analysis
- Application
- Comprehension
- Knowledge

Lower Order Thinking Skills

2001

Higher Order Thinking skills
- Creating
- Evaluating
- Analysing
- Applying
- Understanding
- Remembering

Lower Order Thinking Skills
<table>
<thead>
<tr>
<th>Levels of Thinking</th>
<th>Possible Tasks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Remembering</td>
<td>Describe where Goldilocks lived</td>
</tr>
<tr>
<td>Understanding</td>
<td>Summarise what the Goldilocks story was about</td>
</tr>
<tr>
<td>Applying</td>
<td>Construct a theory why she went into the house</td>
</tr>
<tr>
<td>Analysing</td>
<td>Differentiate between how she reacted and how you would react</td>
</tr>
<tr>
<td>Evaluating</td>
<td>Assess whether or not you think this really happened to Goldilocks</td>
</tr>
<tr>
<td>Creating</td>
<td>Compose a song, skit or poem to convey the Goldilocks story in a new form</td>
</tr>
</tbody>
</table>
## Example of Bloom’s Taxonomy 2

<table>
<thead>
<tr>
<th>Levels of Thinking</th>
<th>Possible Tasks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Remembering</td>
<td>Define respiration. List the function of respiration.</td>
</tr>
<tr>
<td>Understanding</td>
<td>Describe the exchange of oxygen and carbon dioxide in the lungs and in the cells</td>
</tr>
<tr>
<td>Applying</td>
<td>Compare and illustrate the movement of air into and out of the lungs and in the cells through a flow chart</td>
</tr>
<tr>
<td>Analysing</td>
<td>Describe problems of the respiratory system and identify what makes the upper respiratory system so vulnerable to infection.</td>
</tr>
<tr>
<td>Evaluating</td>
<td>Appraise the impact of healthy living habits on the respiratory system.</td>
</tr>
<tr>
<td>Creating</td>
<td>Devise a plan suitable to your living that will enable you to have a healthier respiratory system.</td>
</tr>
</tbody>
</table>
Bigg’s (SOLO) Taxonomy

As learning progresses it becomes more cognitively demanding.

SOLO, stands for the Structure of the Observed Learning Outcome, is a means of classifying learning outcomes in terms of their complexity, which then enables the assessment of students’ work in terms of its quality not of how many components that they have got right.

In the beginning of the learning, the student provides one or few aspects of the task (unistuctural); those that are further along the continuum provide several aspects, but they tend to be unrelated (multistuctural); further along, the students show that they can integrate these aspects into a whole (relational); and, finally, the ones who are furthest along the continuum are able to generalise that whole to as yet untaught applications (extended abstract).

The diagram lists verbs typical of each such level.
## SOLO Taxonomy

<table>
<thead>
<tr>
<th>Levels of Thinking</th>
<th>Example Verbs</th>
<th>Possible question starters</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>S</strong> Single</td>
<td>• Identify</td>
<td>• Name one thing</td>
</tr>
<tr>
<td></td>
<td>• Follow a procedure</td>
<td>• Identify</td>
</tr>
<tr>
<td></td>
<td>• Recite</td>
<td></td>
</tr>
<tr>
<td><strong>O</strong> Organised</td>
<td>• Classify</td>
<td>• Name two or more things</td>
</tr>
<tr>
<td></td>
<td>• Combine</td>
<td>• List</td>
</tr>
<tr>
<td></td>
<td>• Enumerate</td>
<td>• Combine</td>
</tr>
<tr>
<td><strong>L</strong> Linked</td>
<td>• Relate</td>
<td>• What do the following have in common?</td>
</tr>
<tr>
<td></td>
<td>• Compare</td>
<td>• Compare and contrast</td>
</tr>
<tr>
<td></td>
<td>• Analyse</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Explain Thinking</td>
<td></td>
</tr>
<tr>
<td><strong>O</strong> Organised (multi)</td>
<td>• Generate</td>
<td>• What principle applies here?</td>
</tr>
<tr>
<td></td>
<td>• Hypothesise</td>
<td>• Predict.</td>
</tr>
<tr>
<td></td>
<td>• Theorise</td>
<td>• Imagine</td>
</tr>
</tbody>
</table>
Requirements for writing good HOTS items

Developing multiple-choice items to assess Higher Order Thinking Skills (HOTS) requires

1. a great deal of thought and time.

1. a thorough knowledge of the subject matter - including knowledge of popular fallacies and misconceptions.

2. an understanding of the content standards of the curriculum and student performance.

4. a good understanding of HOTS and what it means to grow in the generic cross-curricular skills.

5. good communication skills.

6. sound technical item writing skills.

7. imagination and ingenuity.
Some hints for constructing questions to assess Higher Order Thinking Skills (HOTS)

1. Write items where the “load” is on the problem to solve rather than on the content. Generally if content is needed to respond to the item that is more difficult than the problem embedded in the item, it should be provided as part of the item.

2. Write items that require students to predict the outcome of a situation rather than simply labelling or listing the specific elements associated with a response.

3. Give examples, and ask for the principle or theory they illustrate.

4. Design items that permit multiple interpretations or solutions; this could require for constructed response items students being asked to justify or explain their answers.

5. Design constructed response or performance items in which students have to explain their reasoning and the items have rubrics that reward higher order thinking.
Quick Quiz (True or False or Don’t Know [Need More Information])

1. Changing the verb in the item can change the cognitive complexity of the item and may require revisions to the answer options.

2. Items that require analysis are the lowest cognitive level.

3. An item that requires students to recall information has a high cognitive and difficulty level.

4. The application of knowledge requires students to use or organise information.

5. There are several models/tools that can be used to organise cognitive levels by the complexity of the thought processes required.
Quick Quiz (True or False or Don’t Know [Need More Information])

6. The cognitive complexity of an item is determined by the complexity of the thought processes that a student uses to answer the question.

7. The higher the cognitive level of the item then the higher the difficulty level.
Some examples of questions assessing more cognitively demanding items

1. Analysing

1. (Science) A scientist performed an epidemiological study to investigate the cause and effect relationship of smoking and lung cancer as follows:

(a) Handed out a scientifically valid questionnaire to all colleagues (n = 144) at work.
(b) Checked that there were an equal number of male and female respondents.
(c) Discovered that there were more non-smoking respondents than smoking respondents. Removed some of the non-smokers until both groups had equal numbers.
(d) Checked that all the respondents had a medical check-up in the past year.
(e) Analysed data, wrote the paper and published it in a scientific blog.

From the information provided, analyse the methodology used by this scientist
Some examples of questions assessing more cognitively demanding items

1. **Analysing**
   2. (Social Science) Analyse the effects of the introduction of railways in different countries in the world.

   3. (English) Read aloud the description of ‘A Beautiful’ man (*Chapter: 7 A visit to Cambridge Honeydew*). What is the most beautiful sentence in the description? Justify your choice.

2. **Evaluating**
   1. (Science) Justify your choice of equipment or resources to perform a first-hand investigation to draw a longitudinal section of xylem tissue.

   2. (Social Science) Assess the influence of the ideologies of communism and capitalism on the origins and development of the Cold War to 1968.
2. Evaluating

3. (English) What did the author mean by saying "Before you, like a lantern whose walls are worn so thin you glimpse only the light inside, is the incandescence of a man. The body, almost irrelevant, exists only like a case made of shadows'. Do you agree or disagree with author? Give reasons for your decision. (Chapter: 7 A visit to Cambridge Honeydew)

3. Creating

1. (Science) Develop a safe work practice procedure for using a piece of chemical application equipment. In your answer, include at least five steps in a logical sequence and explain why they are a significant component of your solution.

2. (Social Science) Write the script of a street play giving the importance of tree plantation and try to enact it in your locality.
Assumptions underlying educational assessments

- Items and tasks are designed to give the student a chance to provide evidence as to how much of the construct they have.
- There are many ways to measure constructs they all have their own strengths and weaknesses: it is important for test users to be aware of the limitations of their assessment procedures.
- Multiple sources of information should be used to measure constructs: important decisions should not be based on the single administration of an assessment procedure.
- Performance on a test or task should be generalisable to non-test behaviours.
- Assessment should provide information that can be used to improve learning.
- Assessments can be conducted in a fair manner: this not only includes the test itself, but also the interpretation of the test scores.
Reflections on session 1

1. Which criticisms of large-scale testing (if any) do you consider to be most serious (tests create anxiety; tests categorise and label students; tests damage student self-concept; tests create self-fulfilling prophecies; etc.) in Bhutan? What steps would you take to take account of these criticisms?

2. What do you consider to be the potential uses of computer-based, online testing for large-scale assessments? What are the potential advantages and disadvantages of these uses for your country?