Day 1 Session 13:00 to 14:30

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

In this session we will

• outline the steps involved in designing and developing large scale assessments.
• provide an overall structure for the workshop
• define educational assessment.
• introduce the developmental continuum.
• review the assumptions that underlie educational assessments.
• articulate the purposes of educational assessments.
• discuss the difference between item difficulty and cognition.
Overall structure for designing and developing large scale assessments
Structural overview of the design of large scale assessments

Steps involved in designing and developing large-scale assessments

1. Define the construct to be measured.
2. Develop an assessment framework that shows growth.
3. Build a Table of Specifications.
5. Review the items.
6. Construct field trial assessments.
7. Administer field trial assessments.
8. Analyse field trial results.
9. Select final assessments
10. Administer final assessments
11. Report results
Structure of workshop
Structural overview of the workshop

Steps involved in designing and developing large-scale assessments

1. Conceptual model for developing measurement scales for large scale assessments (introduction to modern assessment theory).
2. Key principles underpinning assessment (validity and reliability).
5. Review assessment items.
6. Construct field trials.
7. Analysis of items and assessments.
8. Select final assessments.
9. Analysis of items and prepare reports.
Definition of educational assessment
Definition of assessment

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

Unstructured
- chance meetings
- conversations

Slightly structured
- questionnaires
- observation
- student self-assessment

More structured
- classroom tests
- checklists
- practical work
- project work
- case studies

Most structured
- examinations
- standardised tests
- published aptitude tests
Developmental continuum
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.
Example developmental continuum for recitation (Grades I & II)

- Can recite a poem or a story with proper speed, diction, expression and tone.
- Can recite a poem or a story with proper speed or expression but makes occasional mistakes in pronunciation or forgets at times.
- Can recite a poem, or story with occasional prompting. Expression is not very strong and effective.
- Cannot recite an entire poem or story without prompting. Pronunciation expression is not appropriate.
- Recitation is poor. Lacks expression.
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 =
Developmental continuum for showing GROWTH in an area of learning.
Capacity Development Workshop

Design and Development of Large-Scale Learning Assessments

Diagram:
- Health & Phys.Ed.
- Science
- The Arts
- English
- Mathematics
- Languages
- Studies of Society
- Technology
Capacity Development Workshop

Design and Development of Large-Scale Learning Assessments
Developmental continuum for showing GROWTH in an area of learning.
Developmental continuum for the New South Wales HSC
Band 6 Biology Description

The typical performance in this band:

Band 6

- demonstrates an extensive and detailed knowledge and superior understanding of biological concepts, including complex and abstract ideas
- demonstrates an extensive understanding of the historical development of biological concepts, their applications and implications for society and the environment, and the future directions of biological research
- communicates succinctly, logically and sequentially using a variety of scientific formats, including diagrams, graphs, tables, flow charts and equations relating to biology
- analyses and evaluates data effectively, identifying biological relationships, quantifying explanations and descriptions, synthesising information to draw conclusions
- uses precise biological terms extensively and correctly in a wide range of contexts
- designs valid experimental processes using appropriate technologies and incorporating the thorough knowledge of the use of a control, variables and repetition to solve biological problems
- applies knowledge and information to unfamiliar situations and designs an original solution to a biological problem
Band 5 and 4 Biology Description

The typical performance in this band:

Band 5
- demonstrates thorough knowledge and understanding of most biological concepts
- demonstrates a thorough understanding of the historical development of biological concepts and their applications and implications for society and the environment
- communicates effectively in a variety of scientific formats including diagrams, graphs, tables, flow charts and equations relating to biology
- explains qualitative and quantitative biological relationships and ideas coherently; identifies patterns in data to draw conclusions
- uses precise biological terms frequently and correctly in a range of contexts
- identifies the correct application of scientific experimental methodology to solve biological problems

Band 4
- demonstrates sound knowledge and clear understanding of some biological concepts
- demonstrates a sound understanding of the historical development of biological concepts and their applications for society and the environment
- communicates using clear written expression and incorporating diagrams of biological structures
- provides qualitative and quantitative descriptions of biological phenomena and explains straightforward biological relationships
- uses general biological terms frequently and correctly in a range of contexts
- identifies the correct components of the experimental scientific method in biology
Performance Achievement Levels or Bands

- Higher level of knowledge and skill
- Band 6
- Band 5
- Band 4
- Band 3
- Band 2
- Band 1

Developmental Continuum
Example Developmental Continuum for Mathematics

Less complex and demanding

More complex and demanding

Addition & Subtraction

Multiplication & Division

Fractions & Decimals

Probability

Learning Objectives
At Level 6 students can conceptualise, generalise and utilise information based on their investigations and modelling of complex problem situations. They can link different information sources and representations and flexibly translate among them. Students at this level are capable of advanced mathematical thinking and reasoning. These students can apply their insight and understandings along with a mastery of symbolic and formal mathematical operations and relationships to develop new approaches and strategies for attacking novel situations. Students at this level can formulate and precisely communicate their actions and reflections regarding their findings, interpretations, arguments and the appropriateness of these to the original situations.

At Level 5 students can develop and work with models for complex situations, identifying constraints and specifying assumptions. They can select, compare and evaluate appropriate problem-solving strategies for dealing with complex problems related to these models. Students at this level can work strategically using broad, well-developed thinking and reasoning skills, appropriate linked representations, symbolic and formal characterisations and insight pertaining to these situations. They can reflect on their actions and formulate and communicate their interpretations and reasoning.
Proficiency scale descriptions for PISA Mathematical Literacy Levels 1 and 2 (2003-2009)

At Level 2 students can interpret and recognise situations in contexts that require no more than direct inference. They can extract relevant information from a single source and make use of a single representational mode. Students at this level can employ basic algorithms, formulae, procedures, or conventions. They are capable of direct reasoning and making literal interpretations of the results.

At Level 1 students can answer questions involving familiar contexts where all relevant information is present and the questions are clearly defined. They are able to identify information and to carry out routine procedures according to direct instructions in explicit situations. They can perform actions that are obvious and follow immediately from the given stimuli.
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 learning objectives. 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.
Example Developmental Continuum Defined by Items for Mathematics

Less complex and demanding

More complex and demanding

Addition & Subtraction

1 2 3 6 10 13

Multiplication & Division

4 5 9 12 15

Fractions & Decimals

7 8 11 16 17 18

Chance

14 19 20
Assumptions underlying educational assessment
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.
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.
Purposes of assessment
Purposes of educational assessments

- **Student assessments**: assessments can generally be used to monitor the performance of students and improve learning. These generally include summative and formative assessments.

- **Instructional decisions**: assessments can be used to provide information that helps teachers monitor their effectiveness and enhance their teaching practices.

- **Selection, placement and certification**.

- **Policy decisions**: assessments are used to evaluate existing policy decisions and provide efficacy evidence for future policy decisions concerning, for example, curriculum and the effectiveness of pedagogical practices.

- **Counselling and guidance decisions**: assessments can be used to help self understanding and help
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

<table>
<thead>
<tr>
<th>Higher Order Thinking skills</th>
<th>Lower Order Thinking Skills</th>
</tr>
</thead>
<tbody>
<tr>
<td>Evaluation</td>
<td>Knowledge</td>
</tr>
<tr>
<td>Synthesis</td>
<td></td>
</tr>
<tr>
<td>Analysis</td>
<td></td>
</tr>
<tr>
<td>Application</td>
<td></td>
</tr>
<tr>
<td>Comprehension</td>
<td></td>
</tr>
</tbody>
</table>

1950

<table>
<thead>
<tr>
<th>Higher Order Thinking skills</th>
<th>Lower Order Thinking Skills</th>
</tr>
</thead>
<tbody>
<tr>
<td>Evaluation</td>
<td>Knowledge</td>
</tr>
<tr>
<td>Synthesis</td>
<td></td>
</tr>
<tr>
<td>Analysis</td>
<td></td>
</tr>
<tr>
<td>Application</td>
<td></td>
</tr>
<tr>
<td>Comprehension</td>
<td></td>
</tr>
</tbody>
</table>

2001

<table>
<thead>
<tr>
<th>Higher Order Thinking skills</th>
<th>Lower Order Thinking Skills</th>
</tr>
</thead>
<tbody>
<tr>
<td>Evaluation</td>
<td>Knowledge</td>
</tr>
<tr>
<td>Synthesis</td>
<td></td>
</tr>
<tr>
<td>Analysis</td>
<td></td>
</tr>
<tr>
<td>Application</td>
<td></td>
</tr>
<tr>
<td>Comprehension</td>
<td></td>
</tr>
</tbody>
</table>
## Example of Bloom’s Taxonomy 1

<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>
Example Developmental Continuum Defined by Items for Mathematics

Less complex and demanding → More complex and demanding

Addition & Subtraction
1 2 3 6 10 13

Multiplication & Division
4 5 9 12 15

Fractions & Decimals
7 8 11 16 17 18

Chance
14 19 20

More cognitively complex or demanding
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 organize information.

5. There are several models/tools that can be used to organize 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.
In-Country Reflections

1. What are some of the issues and challenges involved in constructing assessment frameworks in your country? What have you done to address these issues and challenges?

2. 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 your country? What steps would you take to take account of these criticisms?

3. 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?