Quality pyramid for contemporary assessment

- Quality of items: Objectivity, relevance, efficiency, difficulty level
- Quality of test: Validity, reliability and usability, including the preparation, run and feedback. Coherence between items.
- Quality of the assessment program: a tuned distribution of assessment methods over the whole educational program.
- Quality of lifelong learning: Contribution of assessment to sustainable results of learning, contribution to lifelong learning

Sustainable assessment

- The student central is the basic principle
- Accent on learning for the future
- Meaningful assessment tasks
- Stimulating self-directedness
- Student feels to be the owner of the test
- Sustainability of knowledge is recognized
- Assessment is the start of a new learning cycle
- Clear criteria and standards
- Results of assessment are rich of information
- Informal and non-formal learning is recognized.

Quality of an assessment program*:

- Acceptability
- Authenticity
- Cognitive complexity
- Comparability

Quality of tests:

- Forms of assessment
- Quality criteria for each assessment:
  - Usability
  - Reliability
  - Validity
Forms of assessment

- Paper and pencil tests
- Online tests
- Performance assessment
- Portfolio assessment
- Assessment of prior learning (APL, RPL, …)
- Self assessment
- Peer assessment
- 360° feedback
- …

Criteria

Consequences of a test are high for student: pass or fail
Therefore:

- Usability:
  - Efficiency
  - Fairness
  - Time limit
- Reliability:
  - Consistency
  - Number of questions
  - Standardization
- Validity: measure what is meant to be measured
  - Content validity
  - Balance: sample

Blueprint (test with one item type)

<table>
<thead>
<tr>
<th>Cluster</th>
<th>C1</th>
<th>C2</th>
<th>C3</th>
<th>C4</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leerniveau</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Knowledge</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>5</td>
<td>15</td>
</tr>
<tr>
<td>Insight</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>10</td>
</tr>
<tr>
<td>use</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>6</td>
<td>7</td>
<td>9</td>
<td>8</td>
<td>30</td>
</tr>
</tbody>
</table>

Blueprint (test with one item type)

<table>
<thead>
<tr>
<th>Cluster</th>
<th>C1</th>
<th>C2</th>
<th>C3</th>
<th>C4</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leerniveau</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Knowledge</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Insight</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>use</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>30</td>
</tr>
</tbody>
</table>

Blueprint (test with multiple item types)

<table>
<thead>
<tr>
<th>Cluster</th>
<th>C1</th>
<th>C2</th>
<th>C3</th>
<th>C4</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leerniveau</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Knowledge</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Insight</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Use</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>video</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>6</td>
<td>7</td>
<td>9</td>
<td>8</td>
<td>30</td>
</tr>
</tbody>
</table>
Quality of Items

- Item types (Scalise & Gifford, 2006)
- Quality criteria

Innovation

- The level of innovation depends on
  - The format (closed versus open)
  - The way of construction
  - The type of delivery (paper and pencil versus online)
  - The response tool (keyboard, mouse)
  - Use of multimedia
  - Level of interactivity (feedback)
  - Scoring
  - Evaluation

Test blueprint

Basis for item construction:
- Content: classification of the course
- Level:
  - Knowledge, reproduction
  - Insight, understanding production
  - Use, operationalisation of knowledge
- Item type

Criteria for items

- Relevance
- Objectivity
- Efficiency
- Level of difficulty
- See checklist

A few examples
Relevance

• Describe the history of the Open University.

• Better:
  • Which were the two most important political reasons to start the Open University in Tanzania?

Objectivity

What is the mean size of families?

a 1
b 2
c 3

Better:

In the Netherlands the mean numbers of children per couple of parents in 2009 was near by

a 1.
b 2.
c 3.

Objectivity

What is the definition of authenticity? And which dimensions are distinguished in authenticity? (5 points)

Answer:

Authenticity of an assessment task relates to the degree of resemblance of the task to the future professional life.

Five dimensions are distinguished in authenticity: the assessment task, the physical context, the social context, the assessment result or form, and the assessment criteria.

Efficiency

According the taxonomy for E-learning assessment questions and tasks of Scalise and Gifford (2006), an example of a complex completion item is

a. an item type as an true/false question.
b. an item type as a performance.
c. an item type as a cloze procedure
Better

Given the 'taxonomy for E-learning assessment questions and tasks of Scalise and Gifford (2006)'.

What is an example of a complex completion item?

a. true/false question
b. performance
c. cloze procedure

Level of difficulty

Copenhagen is de capital of

a. Denmark.
b. Finland.
c. Germany.
d. Holland.

Easier:

Copenhagen is the capital of

a. Cambodia.
b. Canada.
c. Denmark.
d. Morocco.

Workshop

• Write down two learning objectives of the course you are involved
• Construct for each learning objective two items, each with a different item type
• Change your items with the items of a colleague
• Screen the items of your colleague on the quality criteria
• Discuss your findings with your colleague

Test and item analyses

Test analyses
• Pass and fail percentage
• Reliability

Item analyses
• Item difficulty (p-value)
• Item discrimination
Questions?

- Desiree.joosten-tenbrinke@ou.nl