DESIGNING LEARNING ASSESSMENTS

SUMMARY NOTE 2:
Designing Academic Assessments

KEY QUESTIONS:

- How do we plan for development of a learning assessment?
- How can we select and adapt items?
- How do we use items to design an assessment tool?
Assessment Blueprint and Planning

Key stages and things to consider

Every journey needs a map. Designing an assessment is no different.

There are many factors that should influence decisions that are made around the design of a learning assessment.

If these things are not carefully considered, the final result may be data that is not aligned to the intended purpose or the intended users.

It may also result in data that is not valid or reliable, and may even result in an assessment that places undue stress on test takers or test administrators.

To avoid this, the questions here should be considered before design begins. Where possible and appropriate consult with assessment users and the results of previous research to ensure that design decisions are evidence based.
Developing Items

**Selected** versus **Constructed Response Items**

An assessment is composed of items, which need to be adapted or developed.

Generally speaking there are two kinds of items that are used in assessments.

- **Selected Response** items allow the test taker to choose from a number of pre-determined responses (i.e. a multiple choice question).
- **Constructed Response** items give the test taker the freedom to develop their own response of a pre-determined length.

The advantages and disadvantages of these different approaches are shown here. What is important is that all of the information contained in the assessment blueprint is used to determine what kind of items will be used. These items can be taken from a previous assessment, or developed for the assessment. In the case of academic assessments they can be developed to align with the curriculum or academic standards.

**Benefits and risks of using constructed response items**

+ Less susceptible to error from guessing
  
  Can more easily access higher order thinking

- Take longer to score
  
  Need for judgement in scoring is a threat to reliability

**Benefits and risks of using selected response items**

+ More reliable for large numbers of test takers
  
  Remove writing skills from influencing results
  
  Can be marked quicker

- Susceptible to guessing
  
  Cannot easily access higher order skills
## Test Design and Assembly

**Bringing it All Together**

Once Items have been selected, an assessment can be designed and assembled.

Before an assessment goes through pilot testing, there are some simple checks that can be done to ensure that it will be useful for its purpose. These **face validity checks** ensure that the assessment is aiming to measure the intended construct, and that there are no other factors that would affect its ability to provide a valid assessment of students’ abilities.

### Face Validity Checklist

<table>
<thead>
<tr>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>The length of the test is appropriate for the use</td>
<td></td>
</tr>
<tr>
<td>The test is clearly laid out and easy to read.</td>
<td></td>
</tr>
<tr>
<td>Items do not cross multiple pages.</td>
<td></td>
</tr>
<tr>
<td>Stimuli, materials and language appear to be suitable for the use and context.</td>
<td></td>
</tr>
<tr>
<td>The test is likely to measure the intended construct and not other constructs.</td>
<td></td>
</tr>
<tr>
<td>The test appears likely to behave similarly for different groups of students.</td>
<td></td>
</tr>
</tbody>
</table>
Piloting and Item analysis

It is essential that before an assessment is used that it is piloted. Piloting allows for statistical analysis of how different items and combinations of items perform. This can be done using Item Response Theory (IRT).

IRT can produce Item Characteristic Curves – such as the graph shown here – for all items piloted or with previous uses. The graph shows the probability that a student will answer the question correctly on the vertical axis (Y-axis). The ability of the student (in the assessed domain) is on the horizontal axis (X-axis). For this item, the lower ability students to the left are less likely to answer the question correctly (at the bottom) than the higher ability students to the right. As students’ ability gets greater, so does their probability of answering the question correctly.

The further to the right that the curve crosses the 0.5 probability line, the more difficult the item is. The steeper the curve, the greater the ability of the item to discriminate between students of differing ability. The difficulty can and should vary within an assessment instrument. The slope of the curve should not get too flat as this indicates that the item does not provide much information about the ability of the student.

Understanding these properties of all items in an assessment can allow the assessment design team to zoom out and think about which combination of items will be most useful.