

Designing Learning Assessments

Summary Note 2:
Designing Academic Assessments

Key Questions:

1. How do we plan for development of a learning assessment?
2. How can we select and adapt items?
3. 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 assessments 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.

- ➔ **Assessment Purpose**
What results do I want? To inform which decisions? To have what impact?
- ➔ **Assessment Users**
Who will use the results of the assessment? What are their needs?
- ➔ **Assessment design process and risks**
How will the assessment be produced in time? Who will do each task?
- ➔ **Defining constructs and content**
What are you measuring? What are the constituent elements?
- ➔ **Assessment specifications**
What is the make-up of the assessment? What are the main parameters?
- ➔ **Test format and question specifications**
What types of questions will be used? In which proportions?
- ➔ **Number of items and time constraints**
How many items will be used? How are they spread across the domain?
What time constraints will be imposed? How long will the test take?
- ➔ **Method of Administration**
1-to-1 or to a group together? Paper-based or on tablets?

Developing Items

Selected versus Constructed Response Items



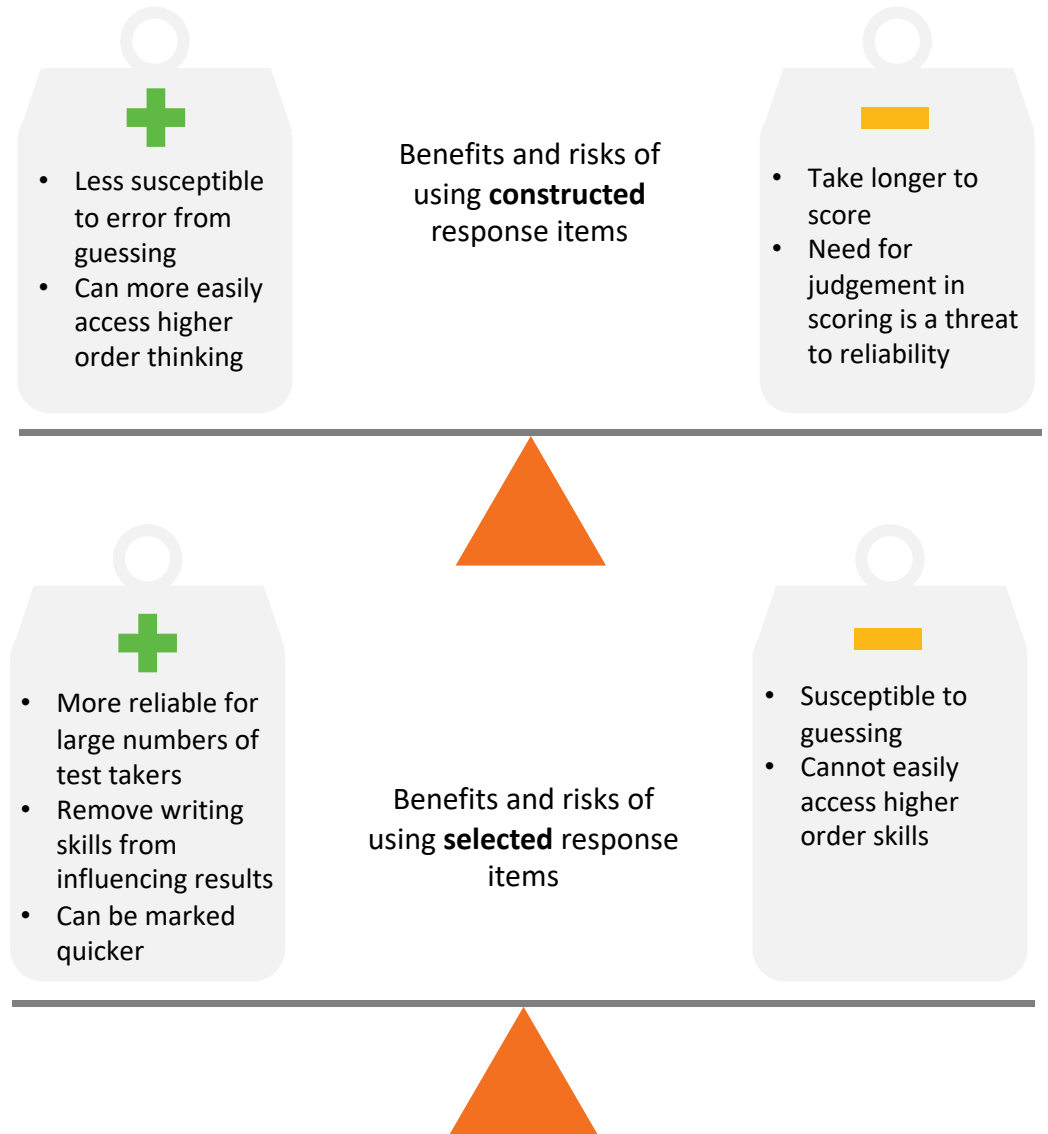
An assessment is composed of items, which need to be adapted or developed.

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.



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.

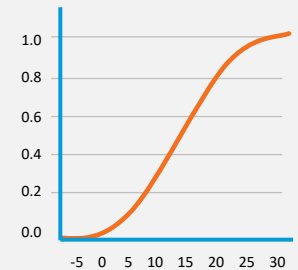


Face Validity Checklist

	Yes	No
The length of the test is appropriate for the use.	<input type="checkbox"/>	<input type="checkbox"/>
The test is clearly laid out and easy to read.	<input type="checkbox"/>	<input type="checkbox"/>
Items do not cross multiple pages.	<input type="checkbox"/>	<input type="checkbox"/>
Stimuli, materials and language appear to be suitable for the use and context.	<input type="checkbox"/>	<input type="checkbox"/>
The test is likely to measure the intended construct and not other constructs.	<input type="checkbox"/>	<input type="checkbox"/>
The test appears likely to behave similarly for different groups of students.	<input type="checkbox"/>	<input type="checkbox"/>

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.