Item panelling, or cognitive walk-through

Ross Turner, ACER
COGNITIVE LABORATORY GUIDELINES DEVELOPMENT

ITEM PANELLING, OR COGNITIVE WALK-THROUGH

Background

Item panelling is a process with a long history at ACER. Essentially the same process is referred to in the US Cognitive Laboratory literature as a ‘cognitive walk-through’. Test development agencies in the UK use the phrase ‘item shredding’ to describe essentially the same process.

Item panelling is one of a number of essential steps in the development of high quality test items. It is a means of subjecting draft test material to the scrutiny of experts who can provide a rigorous evaluation of the quality of the material, and where appropriate can propose ways in which the material may be improved. This is an essential quality assurance process prior to any use of the material with students or other test candidates.

Other quality assurance steps include pre-pilot activities with students, and conducting formal field trials of the material. Each of these procedures has its own special benefits and uses. For example, pre-pilot activities involving students provide a way of validating the outcomes of item panelling, as well as an opportunity to elicit significant information from an audience that matches the eventual test-taking population that might have been overlooked through the item-panelling process. Field trials permit the testing of complete test forms, which generates information about the whole test as well as individual items, and in addition provides the opportunity to gather a range of student responses that can be used, for instance, in the construction of detailed marking guides.

The following guidelines have been developed as a result of test development staff at ACER describing the way they go about panelling items. These guidelines assume that the prior step of articulating a framework for test development and defining the constructs that will be assessed through the test has already been negotiated. That is, a common understanding of the item and test specification for the development task at hand has been established. The client in the project brief should clarify these matters, hopefully in consultation with the test development team and psychometric experts as appropriate. Of course the extent to which these issues have in fact been articulated before the item development task commences varies enormously from project to project.

Let us assume, then, that the following matters have to some extent been established:

- The overall framework for the test (the age(s) or level(s) of students for whom it is intended, the curriculum framework to which it must relate),
- The content of the test (the genres and text types wanted, or the topic areas wanted or any theme to underlie the test),
- Any learning outcomes or other specific skills to be assessed, and

---

1 The author wishes to acknowledge and thank Doug McCurry, and Juliette Mendelovits and her team of test developers, for their contribution to this document. Many of the ideas are theirs.
Any specific constraints regarding item response type, or mix of item response types (eg. Multiple choice, extended answer).

It is common for test development teams to develop a map (often a simple tabulation) of characteristics to be satisfied in the overall test, so that cells in the table can readily be identified and filled in as test development proceeds.

**Guidelines for item development and review**

In essence, the same set of evaluation criteria is appropriate both to item writing and to item review. The following guidelines incorporate a number of these evaluation criteria, and describe a process through which they might be systematically applied in the development of test items.

The cognitive walk-through consists of a small group of expert item writers jointly reviewing material that one or more of them have developed with a view to accepting, modifying or rejecting that material for further development.

Participants will compare their answers to the questions, and raise issues about the questions and the material. It is a robust process of debate that aims to see that the items perform their intended function, are unambiguous and can withstand all kinds of challenges.

**Preparation**

Items should be prepared in a form that is as close to the anticipated production version that is possible given available resources. This includes the presentation of graphic material. The more sensitive the material is to layout issues, the more critical this step is and therefore the closer to final form the material should be. Marking guides, or marking criteria, should be seen as an integral part of an item, and these should be included with each item for consideration by the panel.

There is a limit to the amount of material that can be fruitfully considered in a single review session. This will depend to some extent on the type of material, but a session length of three hours with a break would generally be appropriate.

A cognitive walk-through should involve at least three item-writers and generally not more than five or six.

Material should be distributed before the panel session, to give panel members time to attempt the questions as if each member was a student taking the test, and to explore the material in detail. Panel members should make annotations on their copy of the material both before and during the panel session, and give this annotated material back to the item writer after the session.

All parts of the material should be interrogated in minute detail, including

---

Note, however, that a similar process might also be used to explore less well-formed material, to ‘brainstorm’ ideas at an earlier stage of item development. Clearly the production quality is not such an issue at that stage.
- Any stimulus that forms part of the material, including written stimulus, diagrams or other graphics

- The questions that are posed

- The layout and presentation of the material

- Associated marking guides, and marking criteria.

The following questions will be relevant to an evaluation of the material to a greater or lesser extent, depending on the precise type of material under consideration:

- How does the material relate to the test specification? Where does it fit in the map of requirements for the test?

- Are the questions testing the underlying construct that you are aiming at?

- Is the material interesting? Is it worthwhile? Of some importance?

- Is it at the right level – both in terms of the expected ability level of students, and given the age or year-level(s) of the students to be taking the items?

- Is it coherent? Unambiguous? Clear?

- Is it self-contained? Or does it assume other prior knowledge, and if so is this appropriate?

- Is the question expressed as briefly as possible (particularly for mathematics or numeracy questions, where unnecessarily high reading load should be avoided)?

- Are there any ‘tricks’ in the question that should be removed?

- Is it fair? Are there any ethical matters or other sensitivities that may be breached (for example, racial, ethnic or cultural stereotypes, gender stereotypes, general appropriateness)?

- Do the questions relate to the essence of the stimulus (and vice versa)?

- How will the students perceive this material? To answer this, panel members must imagine the cognitive demands of the items from the point of view of students.

- Is it clear what would constitute an answer to the question? That is, will students know exactly what they are being asked to produce (apart from knowing how)?

- Is the proposed scoring consistent with the underlying ability that is being measured? Would students possessing more of the underlying ability always score better than students with less? For example, in scoring for a mathematics item, are students being inappropriately penalised for poor writing or spelling, or for not showing their working when it is clear they have answered correctly?
− Is it clear how the marking criteria or guidelines would be applied to all possible responses? Are there different approaches to arriving at the same answer? Are all such approaches equally good?

− Does the question require a multi-step approach? Should partial credit be given if part of the answer is achieved?

− When items are grouped as Units, are there dependencies between them? Does one item give a clue to the next one? Is this intentional? Would a different order of items within a Unit make a difference? If a response to one item is incorrect, does this affect possible responses for other items in the Unit?

− How would this material stand up to public scrutiny?

The author of the material should take prime responsibility for recording comments during the panel session. Other comments arising from panel-members' consideration of these questions should be written on copies of the items, and given to the writer of the material at the conclusion of the session.

**The panel session**

Use of a team approach to reviewing items rests on a recognition that more ideas are typically generated through such a process than is the case with individual work.

In the panel session, the group looking at the material should consider essentially the same questions as identified above.

In considering these questions for each item, a decision is required as to whether the item is acceptable as it is, should be modified in ways that deal with flaws or problems identified, or is unacceptable. In many cases consensus among the members of the panel will be a sufficient basis for decision, however the manager responsible for test development will generally have the ultimate decision-making responsibility.

**Revisions, and re-panelling**

Item writers will modify material on the basis of matters raised in the cognitive walkthrough.

The need to subject modified material to a further round of panel scrutiny will depend on the extent of modifications made. This is a matter of judgment. In some cases, an item writer might show a modified version to one member of the panel for advice as to the need for further group scrutiny. All final items are scrutinised by the project director or other appropriate nominated expert before release to clients.