E-assessment: screen design and question difficulty

Changing the way we see test-items in a computer-based environment

Matt Haigh: mrh47@cam.ac.uk

CAMBRIDGE ASSESSMENT

Lessons with a penguin

James is a newly-qualified teacher who is enthusiastic about using technology in the classroom. He has persuaded his head of department to purchase a very impressive software package for assessing his students' science knowledge. The students are particularly excited by the animated penguin that helps them navigate around the features of the software and gives them feedback on their answers.

After several lessons he looks at the students' assessment results. Following a promising rise in their scores, he notices a sudden downturn in performance; even his brightest students are performing badly on the assessments. He checks the software and observes his students more closely in the following lesson. His observations draw a blank and, following a string of incorrect answers, he asks one of his students what is going on:

"Well sir, you see, when you get a question right, the penguin just gives you a little smile, but when you get a question wrong, he rolls off his iceberg and makes a big splash in the water which is really funny, so I'm kind of getting the questions wrong so I can see the penguin get wet"

Background

Aim

The aim of this study was to investigate how aspects of screen design in a computer based assessment affect the difficulty of the test items.

Research questions

1. What are the effects of changing the student-item interaction on measures of item difficulty with a computer-based test item?
2. What insights are given by the test users for any potential sources of difficulty in computer-based test items?

Context and sample

The research was undertaken with 112 15-year-old students in seven English secondary schools. The computer-based assessment was based on GCSE Science qualifications and was administered online.

Research design

Mixed methods

The study used a mixed methods design, with an initial quantitative phase to examine measures of item difficulty, followed by a qualitative phase to collect the perceptions of difficulty from the students taking the assessment.

Stage 1 – Quantitative

- Two parallel forms of the online test were created; in the second of these forms a specific aspect of screen design was modified.
- Students taking the test were randomly assigned one of the two parallel forms.
- The test consisted of five items common to both tests for the purposes of checking that the samples were matched with regards their ability.
- Ten further items were modified for each parallel form.
- Student responses were analysed to determine the measures of item difficulty.
- Background data were collected on prior attainment and ICT competence which have been shown to be key factors in computer-based test performance.

Stage 2 - Qualitative

- Two focus groups (each with four students) were held following the online-test.
- Students were presented with the parallel forms of each item.
- Responses were collected on features that students perceived to make an item easier or more difficult.
- Responses were analysed for common themes to identify insights into how the screen design altered their perception of an item's difficulty.

Results

Analysis and implications

Measures of difficulty

No significant differences were reported in the measures of item difficulty in any of the ten items modified in the test. This implies that students are able to demonstrate their subject knowledge through a variety of screen formats and interactions.

Student perceptions

In many cases students clearly articulate perceived differences in the difficulty of items. On some they are able to reach consensus about what makes an item more difficult, and on others they list factors relating to both versions of the item that have an influence on difficulty.

Implications for practice

The quantitative data indicates that students are able to handle a range of interactions for providing evidence of their abilities which is reassuring for the development of tests in different formats. However the qualitative data indicate that some forms are perceived differently by students in terms of their difficulty and that practitioners should take note of student input when designing computer-based assessments.

Further research

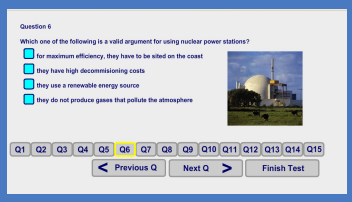
It may be that particular sub-groups of students (e.g. those with poor ICT skills or of lower ability) may be more susceptible to changes in the presentation of items on screen and this is worth investigating further. The study was undertaken with a limited number of screen-modifications and in a specific context. Further research could be undertaken with a broader range of item types and in a wider set of contexts to identify whether the findings were more generalisable.

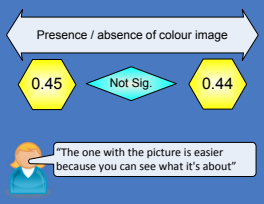
Parallel test forms – Spot the difference...

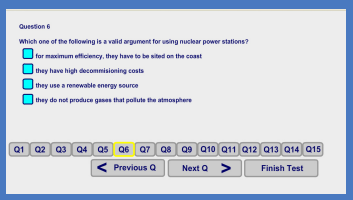
Key to results:

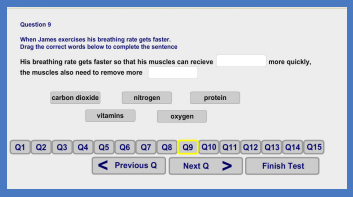
Description of the modification to screen design in the parallel forms of the test	0.55 — The difficulty measure of the question (item facility): 0 represents a very hard question, 1 represents a very easy question
Sig. — Indicates whether the difference in question difficulty is significant (at 5% level)	Student comments selected from the qualitative phase of the study

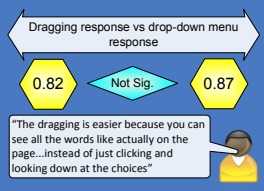
A subset of five out of the ten modified items in the study are shown below

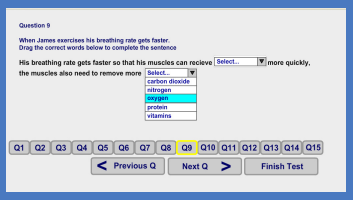


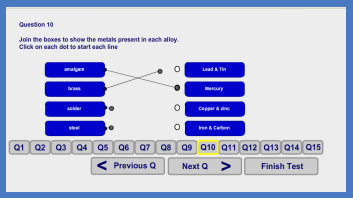


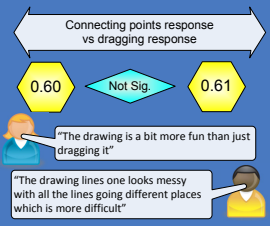


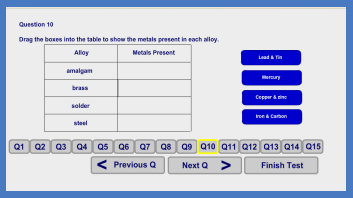


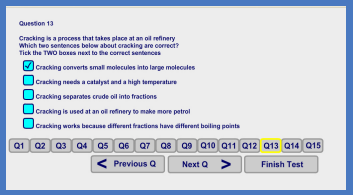


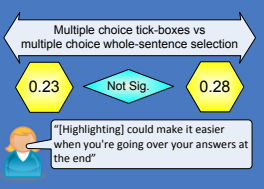


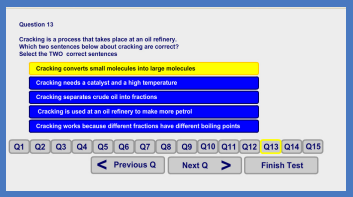


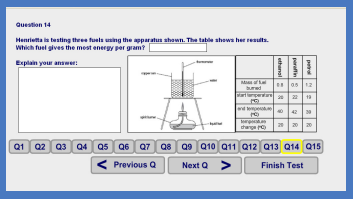


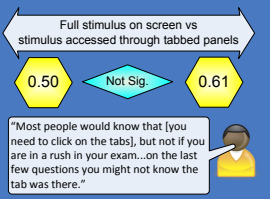


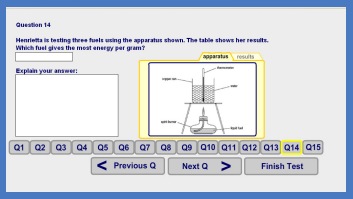












Qualitative analysis – Word clouds

Student responses to the focus groups were divided into two categories: reasons given for features making an item easier and reasons associated with features that make an item more difficult. The word clouds below are based on a word-frequency analysis where font size directly corresponds to the word-frequency.

This study was undertaken as part of a Masters programme in educational research at the Faculty of Education, University of Cambridge. A full version of the Masters thesis along with samples of the online-tests are available from the following website:

www.assessmentresearch.org.uk