

Don't write *more* questions ... write better questions

Once a class of students has taken a Perception test, you can see how well they performed. What too few tutors do is to check how well *the test itself* performed:

- Did it test what you thought it was testing?
- Was it overkill on some topics and a bit thin on others?
- Was the time limit realistic?
- Were the question values (how many marks they were worth) a consistent reflection of how difficult they were for the class?
- Did it produce the spread of marks you expected for this class?

If you have 100 or more results from a test, Enterprise Reporter is capable of analyzing the effectiveness of the questions, allowing the test to be fine tuned so that it delivers a valid score with the minimum number of questions.

Two useful statistics which can be used to fine-tune Perception assessments are Difficulty and Correlation. Once your questions have been used on a large sample of students (possibly over several years), the values for Difficulty and Correlation can be added to the individual questions as tags, and tests can then be built up by selecting questions with particular tag values.

Difficulty or Facility

This is calculated as the average score for the question divided by the maximum achievable score for the question.

- A difficulty of 0.0 means that the candidates found the question to be very hard (no-one got it right)
- A difficulty of 1.0 means that it is very easy (no-one got it wrong).
- 0.5 is ideal (half the candidates got it right), but
- You may want to deliberately set a few questions with difficulties in the range 0.0 – 0.1 to identify the top 10% (to sort out the 1sts from the 2:1s) and so on.

Correlation or Discrimination

The Discrimination is the statistical correlation of the candidate's scores *for this question* and their scores for the whole test, and it ranges from -1.0 to +1.0.

- A high correlation (i.e. close to +1.0) means that the question is measuring the same thing as the test.
- A low correlation means that there is little correlation between test takers getting the question right and getting a good score in the test as a whole.
- A negative correlation indicates that test takers getting this question right generally did *badly* in the test, i.e. the question is a trap for the unwary.



Question Statistics Report

See question analysis from a testing perspective,

[Launch a Question Statistics Report](#)

This is the section of the Reporting software which analyses the questions. Having selected the test, results can be viewed onscreen, or exported to ASCII for subsequent analysis on a spreadsheet. Note that our Perception system will retain results from the 2004-5 session onwards, so that performance from different cohorts of students can be aggregated or compared. If this year's students had difficulty with the same questions as last year's did, maybe it's time to review your teaching of that topic, or to check whether these questions are testing the right things.

The onscreen question report looks like this:

Question description	Topic	Times presented	Times answered	Maximum score	Mean score	Standard deviation of score	Difficulty	Correlation
01 Do insects have 6 legs?	Sample questions	599	586	1	0.418	0.494	0.418	0.343

A Case Study

All users of the Perception system are invited to take the Tryout test before they use the system for real. In the analysis below, the Tryout test was started 1384 times, but only submitted for marking 480 times. The mean score for the test was 44.3%, and the mean time taken was 6 minutes 57 seconds.

Question Statistics for 480 attempts at the Tryout test				
Answered by	Difficulty or Facility	Discrimination or Correlation	Question	comments
97.7%	.414	+0.367	Insects	Ought to be very easy
91.2%	.114	+0.245	The Untouchables	
92.7%	.238	+0.282	Baseball	
92.7%	.601	+0.416	Cities	
73.5%	.594	+0.34	Pulp Fiction	Those who did it, found it easy.
79.1%	.646	+0.351	Queen's surname	Ought to be very easy
93.7%	.977	+0.187	Gloves on hands	Too easy – omit this question
91.4%	.71	+0.286	Likert	Self-answering question
91.4%	.162	+0.312	James Bond	Difficult to get exactly right
82.3%	.069	+0.353	Prime Ministers	Really difficult
83.7%	.111	+0.277	Taco hotspot	Students can't find refectories!
85%	.655	+0.48	Library hotspot	But they know where the library is!
83.1%	.374	+0.428	Sports hotspot	
83.1%	.614	+0.269	Cuckoo	
84.8%	.462	+0.444	Splice joint	
92.3%	.573	+0.564	Tacoma Narrows	High correlation due to endurance?

The test was optional, so few candidates answered every question. The Pulp Fiction question illustrates this – the null responses do not show up in the statistics, and clearly the 26.5% of people who didn't answer it thought they would get it wrong, so they simply missed it out. So null responses can be important.

The way the question template is set up, it is impossible to submit your answers until you have viewed all of the questions. That typically leaves the last question in view while you decide whether to submit your answers, and may account for the fact that the last question was attempted more often than those immediately before it. And it turns out that this was a quite difficult question where the score correlated well with the overall score for the test – indicating that it is one to keep for future use.

The highest correlations (+0.4 and above) mostly came with the later questions. Since this was an optional test, only the candidates who are good (or who think they are) will have got as far as the later questions.

Conclusions

Question statistics taken from a large sample of tests provide useful information on the relevance and effectiveness of your questions. It is better to ask a few incisive, robust questions than adopt a 'shotgun' approach.

As CAA usage develops to contribute an increasing proportion of a student's module marks, tutors and students need assurance that the testing methods are working as they are designed to do. We recommend that an examination of the question statistics for CAA tests be carried out at the end of the semester in which the test was set.

For concise and readable hints about designing better questions, see "[Designing Objective Questions](#)".