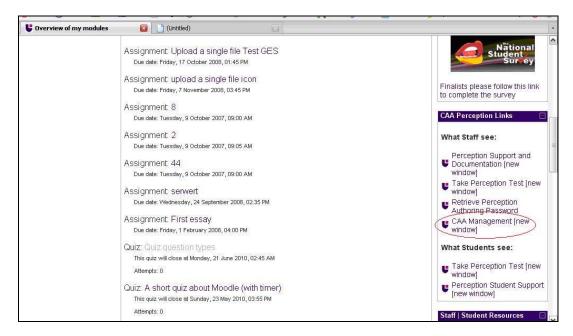# Perception Reports

One of the advantages of using CAA is that a variety of analysis can be run automatically to determine how well individual questions are performing. Statistical information enables the lecturer to improve or eliminate weak questions. Additionally the performance of test-takers can be analysed and this information can help the lecturer to identify:

- General problem areas within the course or module
- Students who are struggling
- Students who are perhaps not being sufficiently challenged
- Different cohorts of students

The CAA Management function on your MyModules page gives you access to the Report area. Results can be seen for any tests that have been taken by students on modules for which you are the internal examiner.

If you want to examine the marks for other modules you need to get the internal examiner of that module to add you to the module's staff list on Learn. This might apply, for example, if you wanted last year's results for a module you have just taken over, or results for a module you teach on but are not internal examiner for. This is a security measure: staff not involved in a module, and all students, are denied access to CAA results.

The Enterprise Reporter index page lists a bewildering array of types of report both test focussed and student focussed. Refer to the write better questions document to find out how some of the reporting statistics can help you.

Assessment Overview Report
Grade Book Report
Question Statistics Report
Survey Report
Coaching Report
Gap Report
Item Analysis Report
Score List Report
Transcript Report

Assessment Overview Report

An Assessment Overview report gives some statistics about the number of times each assessment has been started and finished, and the average scores and times taken. It presents statistics in the form of a histogram.

What information can we obtain from a histogram?

Histograms are good way of presenting the spread of data, in this instance it is an effective, practical working tool in the early stages of data analysis. They are also useful for displaying easily the variation in the process (e.g. Which topics are causing the most difficulty for students?) Depending on the distribution of the data, histograms can be interpreted in various ways.

Considerations:

- A significant number of results should be collected in order for the histogram to show a true likeness of behaviour.
- Questions and assessments will need to be set up accurately in order to gain the maximum benefits of histogram information.
- Actual data values cannot be determined from a histogram, only the number of occurrences within intervals.

Grade Book Report

The Grade Book report displays a table of participants and assessments giving the score for each participant at each assessment. The report can also show average and total weighted scores. It's useful if your students take several assessments and you want to present a report with the results of all of them.

Considerations:

- Grade Book must be defined before this report can be run (Report Manager function required to set this up).
- If participants make more than one attempt at the assessment, only one of the assessment results will be displayed (the assessment displayed depends on which option you select).
- Is it worthwhile just to see a table of results for multiple assessments, this information can be obtained from other reports.

Question Statistics Report

This report largely focuses on question analysis rather than result analysis. It would make a useful tool to look at question validity, efficiency and to determine good quality questions based on the information that it provides.

The report contents itself are defined by a **question detail information** section, and an **outcome analysis information**. The former describes information displayed for each question and the latter describes information displayed for each outcome within a question.

Within the question detail information, the report displays the **difficulty** and **correlation (or discrimination)** of questions.

Difficulty

The difficulty refers to how difficult or easy a question is. It is measured on a scale of 0 to 1 where 0 represents difficult and 1 represents easy.

The following formula can be used to determine the difficulty of a question:

$$\text{Diff}(X) = \frac{\bar{X}}{X\max}$$

where:

Diff(X) = the difficulty of question X

$\overline{X}$ = the mean mark obtained by all candidates attempting question X
X max = the maximum mark available on the question.

Correlation / Discrimination

Correlation measures whether there is any relationship in getting a question right and obtaining a good score in the assessment and it is measured on a scale of -1 (low correlation i.e. little relationship) to 1 (high correlation i.e. measures the same thing as the assessment). For example, if there was a question on football in a maths assessment then this question would have a low correlation. Participants who get it right are no more likely to do well then those who get it wrong. It helps to measure the power of an individual question to discriminate between the stronger and weaker students, it does so by looking at the correlation between the performance of a participant against their overall score on the assessment. The formulae is based on the Pearson product-moment correlation:

$$rxy = \frac{\sum xy}{Ns_x S_y}$$

where:

rxy = the correlation between the item (x) and the test total (y)
xy = the sum of the products of the deviations of the items and the totals
N = the number of observations
sx = the standard deviation of the item
Sy = the standard deviation of the total marks

Considerations:

- Difficulty can be a useful measure of how effective your questions are, but may not be meaningful if you only have a small number of results for the question.
- If you are assessing to differentiate between people, questions are most useful if they are neither too easy nor too hard. A very easy question is answered correctly by everyone, and so doesn't help you measure skills. Similarly a very hard question is answered correctly by very few people, and so adds little to your measurement. An ideal difficulty is 0.5, which roughly means that half your participants answered the question correctly. You may want to look carefully at questions with difficulty of more than 0.75 or less than 0.25, as these questions are less good at discriminating between people.
- If you set a competency test that you expect almost everyone to get 90% or more in, then any question with a difficulty of 0.5 or less may be worth checking to see that it is not incorrectly set or ambiguous.
- Questions with a correlation of less than 0.25, or with a negative correlation, indicate that participants who answer them well do not necessarily do well in the assessment as a whole. Consequently you may wish to consider removing them from the assessment, unless there are other reasons for having them in.
- **Mean for outcome** can be a very useful indicator – for example if people who choose an incorrect outcome score highly in the assessment, then the question might be poorly worded.

- The usefulness of the outcome analysis depends on how you set up your outcomes when creating the question. You should make an outcome for each different possibility that you want to report on.
- When analysing multiple choice questions you can see how well (or poorly) your distracters are performing through the outcome analysis statistics. **"Distracters should differ from the key in a substantial way, not just in some minor nuance of phrasing or emphasis." (Isaacs 1994)**

Survey Report

Shows an analysis of questions for a single assessment, in the form of graphs. This report focuses more on the answers rather than questions.

Considerations:

- To obtain meaningful information from the graphs, requires you to set up the outcomes usefully when designing the question.
- List answers in report could lead to very large reports if many people have answered your questions and this option is only really suitable for open ended question types.

Coaching Report

This is a student focused report which displays, for each participant, a total score, answers given for each question and the score achieved for the question.

Considerations:

- Useful for personal tutoring and identifying weaknesses.
- For each separate attempt that a participant makes, a separate row is shown. It would be more beneficial if the attempts were aggregated together to see whether performance is improved over the number of attempts i.e. if students are learning from previous attempts or if they are blindly taking assessments.

Gap Report

The Gap report lets you compare two sets of results with each other whether it be comparing two groups or two individuals. The differences are shown numerically and graphically. The report can also calculate statistical measures of the difference and the report can include a graphical representation of the difference, so you can easily spot anomalies.

Considerations:

- Compare the results of students who take a pre-test at the start of a course and a post-test after it, and see how much improvement there is from the exercise.
- Compare different classes or categories of learners, and look at differences between their results.
- Conduct needs analysis to justify extra tuition. For example, you get a high achiever student to answer a test, and then compare the results of students against this, to see in which areas they need further tutoring.
- Examine the range of scores to a assessment to help set a cut off or pass score.
- There is the option to choose two different assessments to compare against. This only makes sense if the two assessments are designed to measure the same thing, or have topics in common.

Item Analysis Report

This report is similar to the question statistics report in that they both measure the performance of questions in terms of difficulty and relevance. This report uses the method of classical test construction which is based on psychological assumptions about candidate ability. It provides a simple method of evaluating a test.

There are also histograms showing the distributions of the discriminations and correlations for different questions within the session. By examining the histogram, you can see at a glance how many questions fall into each category.

Considerations

- The statistical data generated allows an evaluation of the quality of the questions, and may point to potential improvements.
- Easy to identify distracters which are implausible and are weakening the question.
- Can identify common areas of student weakness.
- A multiple attempt test is likely to show a higher correct response rate than the same question used in a single attempt assessment.
- The ability of the participant is determined by the difficulty of the test. Therefore, if a test is difficult, participants appear to have a lower ability than when a test is easy.
- The features of particular questions are determined by a group of students taking the test. This makes it problematic to construct tests for a wide group of mixed ability students.
- Questions which are not answered by the participant are treated as if they are incorrect (i.e. scoring zero). If the proportion of participants not answering any question is 10% or more you may consider removing some results before computing the item analysis statistics to gain a more accurate representation of results.
- P value is measured, which is the number of times a question was answered correctly as a proportion of the people who received the question in the assessment. P values should typically range between .20 to .80 with an average value that may vary depending on the purpose of the exam. A broad range achievement test may have an average difficulty value of .50 while a mastery examination may have an average difficulty of .70 or higher.

Score List Report

For a given test, this report shows the performance of every member of the class. It provides standard result information together with the location of taking the test.

Transcript Report

Filtered by participant, this report provides a summary of all the assessments ever taken.