

## From Questionmark on Item Analysis

- **P Value Proportion Correct**

Also known as the Item Difficulty, this is the proportion of times that it was answered correctly, that is the number of times it was answered correctly as a proportion of the people who received the question in a assessment. This is displayed as a number from 0.00 to 1.00. Thus if 90 participants answer an item correctly out of 100 participants the p value for the item is 0.90. The calculation of this statistic ignores partially correct results; a participant is only considered correct if they score at the maximum score available for the question.

P values that are greater than or equal to .90 are relatively too easy items for the sample population, and values that are less than or equal to .10 are relatively too difficult for the sample population. 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. The potential variability or variance of the item difficulty values is maximized if the average p value across items is .50.

- **Item Discrimination**

This is a number that shows how performance on the item differs or discriminates between a high scoring group (the top 27% by assessment score) and a low scoring group (the bottom 27% by assessment score). This number can vary from -1.0 to +1.0. The higher the discrimination, the better the question discriminates between people who score highly in the assessment as compared to those who score lowly. With other factors being held constant, the maximum discrimination is attained when the p value or difficulty for the question and for the average question in the test is at 0.50.

Item Discrimination	Meaning
Negative	This means that there was a higher proportion of the low scoring group who answered the question correctly than that from the high scoring group. These questions should be investigated to determine possible reasons for the reverse expected difference between high and low scoring groups. Possible reasons could include a mis-keyed correct answer, ambiguities in the question which allow low scoring examinees to answer it correctly, or ambiguities which require the high scoring examinees to puzzle over the question or read more into it than was intended.
0.0 and 0.15	These questions should be reviewed to see if any editorial improvements are needed in the questions to produce greater differences between high and low scoring examinees.
0.16 to 0.29	Moderately discriminating questions
0.30 to .49	Strongly discriminating questions
0.50 upwards	Highly discriminating questions

- **Item Total Correlation**

This is the point biserial correlation between the item score and the test score. A high correlation means that people who do well on the question also do well on the test. The following table illustrates possible meanings for this value:

Item Total Correlation	Meaning
Negative	This means that higher scoring participants answered the question incorrectly while lower scoring participants answered it correctly. The question probably needs removal or review.
Around 0.0	No predictive relationship between the question score and the test score. The question probably needs removal or review.
0.0 to 0.19	Low correlation to total scores

0.20 to 0.29	Moderately sized correlation to total scores
0.30 to 0.44	Strong correlation to total scores
Greater than 0.45	Very strong correlation to total scores

### Information displayed for each outcome

The Item Analysis report also lists each outcome for the question. Outcomes are created when you author the question. How many outcomes there are and what their meaning is depends on what is defined when authoring the question. In multiple choice questions, there is usually one outcome per answer choice. The following information is displayed for each outcome:

- The **Outcome Name** is the name or ID of the outcome. A row is also inserted for people who have **Not Answered** the question and for the **TOTAL** number of results by outcome. Outcome names are set in Question Manager, when you create the question.
- **Times Occurred** is the number of times each outcome occurred. It's possible that in some questions, two or more outcomes can result and in this case, the total number of times the outcomes occurs may not be the same as the number of times the question was asked.
- **Proportion Selected** is the proportion of participants selecting the outcome. This value varies from 0.00 to 1.00. This is sometimes called the P value for the outcome. If the P value for an incorrect outcome is high, this could mean some ambiguity in the choice associated with the outcome, or imply some other reason for investigating the question. And if the P value for an incorrect outcome is zero or close to zero, it may mean that the choice associated with the outcome is implausible.
- **Upper Group** is the proportion of people in the top 27% by assessment score who choose this outcome. For example, a value of 0.75 would mean that three quarters of the top group of participants chose this outcome.
- **Lower Group** is the proportion of people in the bottom 27% by assessment score who choose this outcome. For example, a value of 0.50 would mean that half the bottom group of participants chose this outcome.
- **Outcome Discrimination** is a number from -1.0 to +1.0 that shows how performance by people in the high scoring and low scoring group varies or discriminates according to whether or not they choose this outcome.

Correct outcomes should have positive outcome discrimination values and incorrect outcomes should have negative outcome discrimination values or values which are close to zero. If the outcome discrimination value is zero this indicates that the outcome did not provide any information for distinguishing or discriminating between the upper and lower scoring groups. If a question is mis-keyed or has some ambiguous elements which are identified by the higher scoring group but not by the lower scoring group, the outcome discrimination for one or more of the incorrect answer outcomes can be higher than the discrimination value for the correct outcome.

- The **Outcome Correlation** is a number from -1.0 to +1.0 that shows the point biserial correlation between choosing this outcome and the assessment score. If there is a single correct outcome, the outcome correlation for this correct outcome should be the same as the correlation for the question as a whole. The correlation for the incorrect outcomes should be negative or considerably lower than the correlation for the correct answer.

### Other information

At the foot of the report, the report shows the average Item Difficulty (P Value), Item Discrimination and Item Total Correlation across the assessment.

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

The following clarifications and notes apply to the data produced by this report.

- The report only applies to finished assessments. Results from unfinished assessments (where the final page was not seen) are ignored.
- The report omits any questions in the assessment which are not scored, for example explanation questions or surveying questions which have a maximum score of 0 points.
- Negative question scores are not meaningful in several of the statistics calculated. If any question has any negative scores, they are treated as zero for the purposes of this report.

- If the maximum score for a question varies amongst different answers in the same assessment, the report will not calculate statistics for this question. If, for example, a question is delivered once, then the question is edited to change the maximum score, and it is delivered a second time, this will result in a different maximum score the second time.