

Canvas provides quiz item analysis statistics for quiz questions in Canvas. The item analysis comma separated values (CSV) file download will help instructors and course designers gauge the effectiveness of their quiz questions. Quiz analysis estimates reliability, difficulty, and discrimination for multiple choice and true/false questions.

Item analysis may not generate results within specific quizzes. Here are a few limitations to consider within analysis reports.

- Currently analysis is designed to work with quizzes that deliver the same questions to all students. Therefore, if you generate the item analysis CSV for quizzes that pull from a question group, the CSV will generate a line representing each version of the quiz.
- The Quiz Item Analysis report currently only supports Multiple Choice and True/False quiz questions.
- If a quiz allows multiple attempts, the calculations will only consider the student's first attempt (to rule out practice effects).
- Questions that are left blank or not answered count as wrong when determining the total score for the exam, but they are excluded from question-level calculations.
- Statistics only include data for students who attempted to take the quiz.

Canvas Quiz item analysis generates scores based on Cronbach's alpha. Cronbach's alpha measures internal consistency of how closely related a set of items are as a group. Canvas generates an alpha score so long as there are two or more questions in the quiz and the test variance is greater than zero. A variance greater than zero implies two or more submissions produce different scores.

**Note:** To maintain optimum course performance *in the Canvas interface*, the maximum values for calculation are 1000 submissions or 100 questions. For instance, a quiz with 200 questions will not generate quiz statistics. However, a quiz with 75 questions will generate quiz statistics until the quiz has reached 1000 attempts. Results greater than these maximum values can be viewed by downloading the Student Analysis report and viewing the CSV file.

## Quiz Analysis Measurements

### Reliability

Reliability is a measure of the test's internal consistency, meaning if several questions are designed to measure the same information, the test-taker will answer them in a similar way. For example, if a test is given to measure enjoyment of ice cream, students who like ice cream should agree with statements such as "I like ice cream" and "I've enjoyed eating ice cream in the past". Those students should also disagree with statements such as "I hate ice cream".

### Difficulty

The difficulty index (also known as a p-value) shows how hard it is to answer the question correctly. The index is computed as the proportion of students who answered correctly. Proportions range between 0 and 1. Canvas makes this calculation with the point biserial. A point biserial is a correlation coefficient that relates observed item responses and is especially used when one set of data is dichotomous, meaning it can take multiple values based on correct and incorrect responses. In addition to the point biserial of the correct answer, the same

calculation is created for the distractors (also known as a distractor efficiency). Ideally, all of the question's incorrect answers should be equally appealing to the students who miss the question. Scores for this range from -1 to 1.

### Discrimination

Quiz statistics for True/False and Multiple Choice quiz questions include an item discrimination index, which attempts to look at a spread of scores and reflect differences in student achievement. This metric provides a measure of how well a single question can tell the difference (or discriminate) between students who do well on an exam and those who do not. It divides students into three groups based on their score on the whole quiz and displays those groups by who answered the question correctly. Student groups are generally divided as the top 27%, the middle 46%, and the bottom 27%. Ideally, students who did well on the exam should get the question right. If students do well on the overall exam but not on the question, the question itself may need to be revised.

Lower discrimination scores are scored +0.24 or lower; good scores are +0.25 or higher. An ideal discrimination index shows students who scored higher on the quiz getting the quiz question right, students who scored lower on the quiz getting the quiz question wrong, and students in the middle range on either side. A discrimination index of zero shows all students getting the quiz question right or wrong.

### CSV Information

The CSV download also provides the following calculations and counts:

- question ID
- question title
- answered student count
- top student count (students in the top 27%)
- middle student count (students in the middle 46%)
- bottom student count (students in the bottom 27%)
- quiz question count (total number of quiz questions)
- correct student count (number of total students who got the answer right)
- wrong student count (number of total students who got the answer wrong)
- correct student ratio (ratio of students who got the answer right)
- wrong student ratio (ratio of students who got the answer wrong)
- correct top student count (students in the top 27% who got the answer right)
- correct middle student count (students in the middle 46% who got the answer right)
- correct bottom student count (students in the bottom 27% who got the answer right)
- variance (of scores on this question)
- standard deviation (of scores on this question)
- difficulty index
- alpha score (for the whole exam)
- point biserial of the correct answer (reliability index)
- point biserial of the first incorrect answer or distractor (followed by the second, etc.)

Users with permission to read SIS data in the course can also view the `sis_id` column in the CSV download.