Improve Testing (and Teaching) with Item Analysis

Did you know that Blackboard’s Test tool includes an Item Analysis Tool that provides statistical data of student performance both by question and overall?

Easily recognize questions that might be poor discriminators of student performance, improve questions for future use, or adjust credit on current attempts. You can run item analyses on deployed tests, but not on surveys.

How to run an Item Analysis

1. Ensure Edit Mode is set to ON
2. Within the contextual menu for the deployed test or in the Grade Center column header, click Item Analysis.
3. The next screen gives you the option to select the test for which you’d like to run the report (A) and to view the report once it has been generated (B).
What information is provided?

After opening the report, you'll notice a statistical summary (A) which includes average score, average time, and question counts broken down by discrimination and difficult calculations. (Descriptions of how these calculations are determined and what they mean can be found below.) The report also includes filters (B) so you can easily see the questions that are deemed "poor discriminators" or "hard" based on student results. Finally, you can see the individual questions with statistical data (C).

Test Summary Statistics - What does the summary page show me?
The summary statistics at the top of the Item Analysis page provide data on the test as a whole:

- **Possible Points** – the total number of points for the test.
- **Possible Questions** – the total number of questions in the test.
- **In Progress Attempts** – the number of students currently taking the test that have not yet submitted it.
- **Completed Attempts** – the number of submitted tests.
- **Average Score** – scores denoted with an * indicate that some attempts are not graded and that the average score might change after all attempts are graded.
- **Average Time** – the average completion time for all submitted attempts.
- **Discrimination** – this area shows the number of questions that fall into Good (greater than 0.3), Fair (between 0.1 and 0.3) and Poor (less than 0.1) categories. A discrimination value is listed as **Cannot Calculate** when the question’s difficulty score is 100% or when all students receive the same score on a question. Questions with discrimination values in the Good and Fair categories are better at differentiating between students with higher and lower levels of knowledge. Questions in the Poor category (less than 0.1) are recommended for review.

	Request Support
ETC@mayo.edu

Monday - Friday
8:00 - 6:00 CST
(907) 286-9087
• **Difficulty** – this area shows number of questions that fall into the **Easy** (greater than 80%), **Medium** (between 30% and 80%) and **Hard** (less than 30%) categories. Difficulty is the percentage of students who answered the question correctly. Questions in the Easy or Hard categories are recommended for review.

Note: Only graded attempts are used in item analysis calculations. If there are attempts in progress, Item Analysis ignores those attempts until they are submitted and you run the item analysis report again.

**Question Statistics Table**

You can filter the question table by question type, discrimination category, and difficulty category.

In general, good questions have:

- **Medium** (30% to 80%) difficulty.
- **Good** or **Fair** (greater than 0.1) discrimination values.

Questions that are recommended for review are indicated with a red circle. They may be of low quality or scored incorrectly. In general, questions recommended for review have:

- **Easy** ( > 80%) or **Hard** ( < 30%) difficulty.
- **Poor** ( < 0.1) discrimination values.

The table provides the following statistics for each question in the test:

- **Discrimination** – indicates how well a question differentiates between students who know the subject matter those who do not. A question is a good discriminator when students who answer the question correctly also do well on the test. Values can range from -1.0 to +1.0 and are calculated using the Pearson Correlation Coefficient. A discrimination value of less than 0.1 or that is negative indicates that the question might need review. Discrimination values cannot be calculated when the question’s difficulty score is 100% or when all students receive the same score on a question.
- **Difficulty** – the percentage of students who answered the question correctly. Difficulty values can range from 0% to 100%, with a high percentage indicating that the question was easy. Questions in the Easy (greater than 80%) or Hard (less than 30%) categories might need review.
- **Graded Attempts** – number of question attempts where grading is complete. Higher numbers of graded attempt produce more reliable calculated statistics.
- **Average Score** – if a question has attempts that need grading, only the graded attempts are used to calculate the average score.
- **Standard Deviation** – measure of how far the scores deviate from the average score. If the scores are tightly grouped, with most of the values being close to the average, the standard deviation is small. If the data set is widely dispersed, with values far from the average, the standard deviation is larger.

Request Support
ETC@mayo.edu

Monday - Friday
8:00 - 5:00 CST
(507) 266-9087
- **Standard Error** – an estimate of the amount of variability in a student’s score due to chance. The smaller the standard error of measurement, the more accurate the measurement provided by the test question.

**Can I drill into a specific question?**

Yes. Click on one of the individual questions to see just which answer was chosen by students, distributed by class quartile.

**What do the individual question detailed statistics mean?**

The following is a provided by Blackboard:

The **Question Details** page displays item analysis statistics for student performance on individual test questions.

Note: Go to help.blackboard.com and search for the term **Item Analysis** to see examples, the mathematical formulas used to calculate Item Analysis statistics, and how the tool handles tests with multiple attempts.
Question Statistics

The top of the **Question Details** page contains statistical data on the question you selected:

- **Discrimination** – indicates how well a question differentiates between students who know the subject matter those who do not. The discrimination score is listed along with its category: **Poor** (less than 0.1), **Fair** (0.1 to 0.3), and **Good** (greater than 0.3). A question is a good discriminator when students who answer the question correctly also do well on the test. Values can range from -1.0 to +1.0 and are calculated using the Pearson Correlation Coefficient. A discrimination value of less than 0.1 or that is negative indicates that the question might need review. Discrimination values cannot be calculated when the question’s difficulty score is 100% or when all students receive the same score on a question.

- **Difficulty** – the percentage of students who answered the question correctly. The difficulty percentage is listed along with its category: **Easy** (greater than 80%), **Medium** (30% to 80%), and **Hard** (less than 30%). Difficulty values can range from 0% to 100%, with a high percentage indicating that the question was easy. Questions that fall in the Easy or Hard categories might need review.

- **Graded Attempts** – number of question attempts where grading is complete. Higher numbers of graded attempt produce more reliable calculated statistics.

- **Average Score** – if a question has attempts that need grading, only the graded attempts are used to calculate the average score.

- **Std Dev** – measure of how far the scores deviate from the average score for this question. If the scores are tightly grouped, with most of the values being close to the average, the standard deviation is small. If the data set is widely dispersed, with values far from the mean, the standard deviation is larger.

- **Std Error** – an estimate of the amount of variability in a student’s score due to chance. The smaller the standard error of measurement, the more accurate the measurement provided by the test question.

- **Skipped** – number of students who skipped this question.
Question

The Question section displays the question text and the answer choices. The information varies depending on the question type.

The following question types list the number of students who selected each answer choice and the distribution of those answers among the class quartiles:

- Multiple Choice
- True/False
- Opinion Scale/Likert

- Multiple Answer
- Either/Or

The following question types list the number of students who selected each answer choice:

- Matching
- Ordering
- Fill in Multiple Blanks

The following question types list the number of students who got the question correct, incorrect, or skipped it:

- Calculated Formula
- Calculated Numeric
- Fill in the Blank
- Hot Spot
- Quiz Bowl

The following question types list the question text only:

- Essay
- File Response
- Short Answer
- Jumbled Sentence (includes list of possible answers)

Answer Distributions

The distribution of answers among the class quartiles is included for Multiple Choice, Multiple Answer, True/False, Either/Or, and Opinion Scale/Likert question types. This shows you the types of students that selected the correct or incorrect answers.

- Top 25%: Number of students with total test scores in the top quarter of the class who selected the answer option.
- 2nd 25%: Number of students with total test scores in the second quarter of the class who selected the answer option.
- 3rd 25%: Number of students with total test scores in the third quarter of the class who selected the answer option.
- Bottom 25%: Number of students with total test scores in the bottom quarter of the class who selected the answer option.