

# How to use the Item Analysis tool in Blackboard 9

## Contents (v1.0)

Overview .....	1
About the Test Summary on the Item Analysis Page .....	4
About the Question Statistics Table on the Item Analysis Page .....	5
How to View Question Details on a Single Question .....	7
Symbol Legend .....	10
Item Analysis and Multiple Attempts, Question Overrides, Question Edits .....	12
Examples .....	12

## Overview

Item analysis provides statistics on overall test performance and individual test questions. This tool contains some of the same information as the Column Statistics and Attempts Statistics discussed earlier, but presents it in a way that helps you quickly evaluate the quality of each question and its ability to discriminate between students who understand the material and those who do not.

You can use item analysis to:

- Improve questions for future test administrations or to adjust credit on current attempts.
- Discuss test results with your class.
- Provide a basis for remedial work.
- Improve classroom instruction.

You can access item analysis in three locations within the assessment workflow. It is available in the contextual menu for a:

- Test deployed in a content area.
- Deployed test listed on the **Tests** page.
- Grade Centre column.

You can run item analyses on deployed tests with submitted attempts, but not on surveys.

**Grade Center : Tests**

Create Column Create Calculated Column Manage Reports

Move To Top Email

Grade Information Bar

Last Name	First Name	Username	Unit 2 Quiz
Brown	Tony	tbrown	30.00
Casper	Chris	ccasper	--
Cooper	Ashby	acooper	30.00
Durand	Porter	pdurand	--

Quick Column Information  
Edit Test  
Grade Attempts  
Grade Anonymously  
Item Analysis  
Attempts Statistics  
Download Results

Status: Complete. Item Analysis of Unit 2 Quiz successfully ran on January 4, 2013 10:47 AM. View Analysis

**Item Analysis : Unit 2 Quiz**

Select Test: Unit 2 Quiz Run

Available Analysis

Unit 2 Quiz - January 4, 2013 12:47 AM

Access previously run item analyses under the **Available Analysis** heading or select a deployed test from the drop-down list and click **Run** to generate a new report. The new report's link appears under the **Available Analysis** heading or in the status receipt at the top of the page.

**NOTE:** For best results, run item analyses on single-attempt tests after all attempts have been submitted and all manually graded questions are scored. Interpret the item analysis data carefully and with the awareness that the statistics are influenced by the number of test attempts, the type of students taking the test, and chance errors.

### QUICK STEPS: How to Run an Item Analysis on a Test

You can run item analyses on tests that include single or multiple attempts, question sets, random blocks, auto-graded question types, and questions that need manual grading. For tests with manually graded questions that have not yet been assigned scores, statistics are generated only for the scored questions. After you manually grade questions, run the item analysis again. Statistics for the manually graded questions are generated and the test summary statistics are then updated.

1. Go to one of the following locations to access item analysis:
  - A test deployed in a content area.
  - A deployed test listed on the **Tests** page.

- A Grade Centre column for a test.
2. Access the test's contextual menu.
  3. Select Item Analysis.
  4. In the **Select Test** drop-down list, select a test. Only deployed tests are listed.
  5. Click **Run**.
  6. View the item analysis by clicking the new report's link under the **Available Analysis** heading or by clicking **View Analysis** in the status receipt at the top of the page.
  7. Click on questions that are recommended for review – indicated by red circles – to investigate if the question is a good discriminator of student knowledge. Questions are flagged for review based on discrimination and difficulty levels.

Filter Questions								
Select Question Type:		Select Discrimination:		Select Difficulty:				
All Question Types		All Discrimination		All Difficulty		Go Reset Filter		
Question	Question Type	Discrimination ▲	Difficulty	Graded Attempts	Average Score	Std Dev	Std Error	
* ● S1: Something you should never put on a bum...	Short Answer (RB)	Cannot Calculate	0.0%	0	0.0	0.0	0.0	
MC2: Which of these has the most medical training...	Multiple Choice (RB)	0.82	50.0%	4	5.0	5.78	2.89	
● O1: Put the following emergency tasks in order...	Ordering	0.83	87.5%	4	8.75	2.5	1.25	

# About the Test Summary on the Item Analysis Page

The Test Summary is located at the top of the item analysis page and provides data on the test as a whole.

**Item Analysis : Chapter 3 Test**

Select Test: Chapter 3 Test [Run](#)

Available Analysis

- Chapter 3 Test - September 4, 2012 4:48 PM
- Chapter 1 Test - September 4, 2012 2:06 PM

**Chapter 3 Test**

Analysis Last Run September 4, 2012 4:48 PM. Run Item Analysis again to display the latest question data [A](#) [Edit Test](#)

**Test Summary** [B](#)

240.0	39	1	8	248.63	00 hr 54 min
Possible Points	Possible Questions	In Progress Attempts	Completed Attempts	Average Score	Average Time

Discrimination

10 Good Questions
4 Fair Questions
5 Poor Questions
20 Cannot Calculate

Difficulty

18 Easy Questions
11 Medium Questions
10 Hard Questions

[A](#) Edit Test provides access to the Test Canvas.

[B](#) The Test Summary provides statistics on the test, including:

**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. The score displayed here is the average score reported for the test in the Grade Centre.

**Average Time** - the average completion time for all submitted attempts.

**Discrimination** - this area shows the number of questions that fall into the **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 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 are recommended for review.

**Difficulty** - this area shows the 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 and are indicated with a red circle.

**NOTE:** Only graded attempts are used in item analysis calculations. If there are attempts in progress, those attempts are ignored until they are submitted and you run the item analysis report again.

## About the Question Statistics Table on the Item Analysis Page

The question statistics table provides item analysis statistics for each question in the test. Questions that are recommended for your review are indicated with red circles so that you can quickly scan for questions that might need revision.

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 red circles. 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.

Filter Questions <span>A</span>							
Select Question Type:		Select Discrimination:		Select Difficulty:		Go Reset Filter	
All Question Types		All Discrimination		All Difficulty			
Question	Question Type	Discrimination <span>C</span>	Difficulty	Graded Attempts	Average Score	Std Dev	Std Error
* <span>●</span> S1: Something you should never put on a burn...	Short Answer (RB)	Cannot Calculate	0.0%	0	0.0	0.0	0.0
MC2: Which of these has the most medical training...	Multiple Choice (RB)	0.82	50.0%	4	5.0	5.78	2.89
<span>●</span> O1: Put the following emergency tasks in order... <span>B</span>	Ordering	0.83	87.5%	4	8.75	2.5	1.25

**A.** Filter the question table by question type, discrimination category, and difficulty.

**B.** Investigate a specific question by clicking its title and reviewing its **Question Details**.

**C.** Statistics for each question are displayed in the table, including:

**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. Questions are flagged for review if their discrimination value is less than 0.1 or is negative. Discrimination values cannot be calculated when the question's difficulty score is 100% or when all students receive the same score on a question.

Discrimination values are calculated with the Pearson correlation coefficient.  $X$  represents the scores of each student on a question and  $Y$  represents the scores of each student on the assessment.

$$r = \frac{1}{n-1} \sum_{i=1}^n \left( \frac{X_i - \bar{X}}{s_X} \right) \left( \frac{Y_i - \bar{Y}}{s_Y} \right)$$

The following variables are the standard score, sample mean, and sample standard deviation, respectively:

$$\frac{X_i - \bar{X}}{s_X}, \bar{X}, \text{ and } s_X$$

**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 are flagged for review.

Difficulty levels that are slightly higher than midway between chance and perfect scores do a better job differentiating students who know the tested material from those who do not. It is important to note that high difficulty values do not assure high levels of discrimination.

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

**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. The score displayed here is the average score reported for the test in the Grade Centre.

**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.

**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.

## How to View Question Details on a Single Question

You can investigate a question that is flagged for your review by accessing its Question Details page. This page displays student performance on the individual test question you selected.

1. On the **Item Analysis** page, scroll down to the question statistics table.
2. Select a linked question title to display the **Question Details** page.

← Return

A

<< < Page 10 of 13 > >>

Chapter 1 Test

Analysis Last Run September 4, 2012 2:06 PM.

B

Edit Test

C	0.82	50.0%	4	5.0	5.78	2.89	0
	Discrimination Good <i>i</i>	Difficulty Medium <i>i</i>	Graded Attempts <i>i</i>	Average Score <i>i</i>	Std Dev <i>i</i>	Std Error <i>i</i>	Skipped <i>i</i>

Multiple Choice(RB): MC2

Question:

• Which of these has the most medical training?

D

Answers	Total <i>i</i>	Top 25% <i>i</i>	2nd 25% <i>i</i>	3rd 25% <i>i</i>	Bottom 25% <i>i</i>
<input type="checkbox"/> EMT	1(25.0%)	0	0	0	1
<input checked="" type="checkbox"/> Paramedic	2(50.0%)	1	0	1	0
<input type="checkbox"/> First Responder	0(0.0%)	0	0	0	0
<input type="checkbox"/> Firefighter	1(25.0%)	0	0	0	1

Legend

• Review recommended

★ Questions might have changed after deployment

\* Not all submissions have been graded

(QS) Question Set

(RB) Random Block

- A. Use the arrows to page through questions sequentially or to skip to the first or last question.
- B. Click **Edit Test** to access the **Test Canvas**.
- C. The summary table displays statistics for the question, including:
  - **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. Questions are flagged for review if their discrimination value is less than 0.1 or is negative. Discrimination values cannot be calculated when the question's difficulty score is 100% or when all students receive the same score on a question.

Discrimination values are calculated with the Pearson correlation coefficient. X represents the scores of each student on a question and Y represents the scores of each student on the assessment.

$$r = \frac{1}{n-1} \sum_{i=1}^n \left( \frac{X_i - \bar{X}}{s_X} \right) \left( \frac{Y_i - \bar{Y}}{s_Y} \right)$$

The following variables are the standard score, sample mean, and sample standard deviation, respectively:

$$\frac{X_i - \bar{X}}{s_X}, \bar{X}, \text{ and } s_X$$

- **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 in the easy or hard categories are flagged for review.

Difficulty levels that are slightly higher than midway between chance and perfect scores do a better job differentiating students who know the tested material from those who do not. It is important to note that high difficulty values do not assure high levels of discrimination.



- **Graded Attempts** - number of question attempts where grading is complete. Higher numbers of graded attempt produce more reliable calculated statistics.
  - **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. The score displayed here is the average score reported for the test in the Grade Centre.
  - **Std Dev** - 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.
  - **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.
- D. The question text and answer choices are displayed. The information varies depending on the question type:

Type of Information Provided	Question Types
Number of students who selected each answer choice  - and -  distribution of those answers among the class quartiles.	Multiple Choice  Multiple Answer  True/False  Either/Or  Opinion Scale/Likert
Number of students who selected each answer choice.	Number of students who selected each answer choice.
Number of students who got the question correct, incorrect, or skipped it.	Calculated Formula  Calculated Numeric  Fill in the Blank  Hot Spot  Quiz Bowl

Type of Information Provided	Question Types
Number of students who got the question correct, incorrect, or skipped it.	Essay File Response Short Answer Jumbled Sentence (also includes the answers students chose from)

### 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. The distribution shows you the types of students who 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.

## Symbol Legend

Symbols appear next to the questions to alert you to possible issues:

Legend	
● Review recommended	▲ Questions might have changed after deployment
* Not all submissions have been graded	(QS) Question Set (RB) Random Block

- **Review recommended:** This condition is triggered when discrimination values are less than 0.1 or when difficulty values are either greater than 80% (question was too easy) or less than 30% (question was too hard). Review the question to determine if it needs revision.
- **Question may have changed after deployment:** Indicates that a part of the question changed since the test was deployed. Changing any part of a question after the test has been deployed could mean that the data for that

question might not be reliable. Attempts submitted after the question was changed may have benefited from the change. This indicator helps you interpret the data with this in mind.

- This indicator is not displayed for restored courses.
- Not all attempts have been graded: Appears for a test containing questions that require manual grading, such as essay questions. In a test containing an essay question with 50 student attempts, this indicator shows until the instructor grades all 50 attempts. The Item Analysis tool uses only attempts that have been graded at the time you run the report.
- (QS) and (RB): Indicate that a question came from a Question Set or Random Block. Due to random question delivery, it is possible that some questions get more attempts than others.

# Item Analysis and Multiple Attempts, Question Overrides, Question Edits

The item analysis tool handles multiple attempts, overrides, and other common scenarios in the following ways:

- When students are allowed to take a test multiple times, the last submitted attempt is used as the input for item analysis. For example, a test allows three attempts and Student A has completed two attempts with a third attempt in progress. Student A's current attempt counts toward the number listed under **In Progress Attempts** and none of Student A's previous attempts are included in the current item analysis data. As soon as Student A submits the third attempt, subsequent item analyses will include Student A's third attempt.
- Grade Centre overrides do not impact the item analysis data because the item analysis tool generates statistical data for questions based on completed student attempts.
- Manually graded questions or changes made to the question text, correct answer choice, partial credit, or points do not update automatically in the item analysis report. Run the analysis again to see if the changes affected the item analysis data.

## Examples

Item analysis can help you improve questions for future test administrations or fix misleading or ambiguous questions in a current test. Some examples are:

- You investigate a Multiple Choice question that was flagged for your review on the item analysis page. More Top 25% students choose answer B, even though A was the correct answer. You realize that the correct answer was miskeyed during question creation. You edit the test question and it is automatically regraded.
- In a multiple choice question, you find that nearly equal numbers of students chose A, B, and C. Examine the answer choices to determine if they were too ambiguous, if the question was too difficult, or if the material was not covered.
- A question is recommended for review because it falls into the hard difficulty category. You examine the question and determine that it is a hard question, but you keep it in the test because it is necessary to adequately test your course objectives.