

# **Quick start guide for using Integrity reports to evaluate and improve the quality of assessments**

<http://integrity.castlerockresearch.com>  
Castle Rock Research Corp.  
November 17, 2005

This guide is designed to provide a very general overview of how Integrity reports can be used to evaluate and improve your assessments. For a more comprehensive guide to interpreting and using results, please see the “Strategies for using Integrity reports to evaluate and improve the quality of assessments” guide found in the **Knowledge base** section of Integrity.

## Using the executive summary report

The three sections of the executive summary report are designed to provide a quick overview of the key findings of the assessment analysis:

- 1) The **Summary section** provides a verbal description of some key aspects of the test and test-item performance
- 2) The **Table of test statistics** provides statistical summary information of the test performance
- 3) The **Histogram of test total scores** provides a graphical representation of the distribution of total test scores (e.g., how many students obtained different scores on the test)

The **Summary section** provides statements that flag potential problems with test items or with the test as a whole, and summarizes key features of the assessment. An example of the **Summary section** of the executive summary is shown in Figure 1. The example shows that item 2 has been flagged as potentially being too difficult an item. Item 2 is also flagged as having a negative discrimination value (CPBR) and, therefore, is possibly being mis-keyed. By clicking on the item number, you will be taken to a detailed item level report where these statements are expanded upon and other characteristics of the item performance are presented.

The summary statements can also be used to obtain more information. For example, one of the summary statements is “The KR-20 for this test indicates moderate to high test reliability.” If you are interested in finding out more about what KR-20 test reliability is and how it relates to your test performance, the next section of the executive summary (**Table of test statistics**) contains that statistic. By clicking on the term within the **Table of test statistics**, you will be presented with the definition of the term, retrieved from the glossary.

**Figure 1.** An example of a summary section within the executive summary produced by Integrity.

Summary	
Summary statements	Applicable items
→ This item may be too difficult. Item difficulty affects item discrimination in that items of high and low difficulty may have lower discrimination statistics. Consider reviewing the content of the item to determine if it should be made less difficult.	2
→ This item has a negative CPBR, which is statistically very problematic. Examinees of low ability may have a greater probability of answering the item correctly, than do examinees of high ability. This item may be keyed incorrectly. If the item key is correct, review the content of the item carefully and consider deleting this item from the test.	2
→ This item has low discrimination. Examinees of low ability should have a much lower probability of answering an item correctly than do examinees of high ability. Low discrimination statistics suggest that this may not be what is occurring. Consider reviewing and revising the content of this item to see if ambiguity in the item content can be limited. Also, consider that item difficulty affects item discrimination in that items of high and low difficulty may have lower discrimination statistics.	3 12 13 14 15 16 17 20 22 23 24 25 36 39 40 41 42 43 45 47 48 49 50 51 52 53 55 56 57 60 61 63 64 65 66 68 70 80 81 83 84 85 86 87 88 90 91 92 94 96 97 98 99
→ This item may be too easy. Item difficulty affects item discrimination in that items of high and low difficulty may have lower discrimination statistics. Consider reviewing the content of the item to determine if it should be made more difficult.	17 42 83
→ The KR-20 for this test indicates moderate to high test reliability.	
→ 3 pairs of examinees have been identified by the collusion detection analysis.	

The **Table of test statistics** also provides information on the maximum and minimum scores achieved on the test, the mean of the test, and other detailed information for the interested user. Information such as the “Maximum score” is useful in telling whether or not the top students in the class obtained the highest possible score on the assessment. If no students obtained the highest possible score (e.g., 55 out of 55), this may be something to investigate further (e.g., pull the exams of the students that received the top score and find out which items they did not get correct). In similar manner, the “Minimum score” is useful in telling us how the struggling students in the class performed on the assessment. All terms in the **Table of test statistics** are clickable and thoroughly defined through the glossary.

The **Histogram of total test scores** provides a visual representation of the number of students achieving each total test score. For example, it shows the number of students who obtained a score of 35 out of 55, 36 out of 55, and so on. This graph is useful in evaluating the distribution of test scores. For example, if the histogram shows that a great number of students received very high scores on the test (i.e., the graph bars are pushed to the left of the graph), this suggests that either the test may have been too easy for the students or that the students mastered the test material at a high level. In any case, further investigation would be appropriate (e.g., reviewing the content of the test items and referring to the item statistics report to evaluate the difficulty of the items – “Are the items less difficult that I would have thought they should be?”).

## Using summary information in the item-level statistical reports

By clicking on an item listed in the executive summary report, you are taken to a detailed item-level statistical report. The item-specific summary statements provide information regarding the performance of that specific item. The summary statements describe characteristics of the items and direct you to the other sections of the item level report. Above the summary statements is an information bar that shows the item number, keyed correct answer to the item, difficulty level of the item, and item discrimination. An example of summary statements and the information bar is shown in Figure 2.

**Figure 2.** An example of summary statements and the information bar from an item-level report produced by Integrity.

<b>Item 51 - Key = B , Difficulty = 0.598 , Discrimination (CPBR) = 0.195</b>
→ This item is of moderate difficulty.
→ This item has low discrimination. Examinees of low ability should have a much lower probability of answering an item correctly than do examinees of high ability. Low discrimination statistics suggest that this may not be what is occurring. Consider reviewing and revising the content of this item to see if ambiguity in the item content can be limited. Also, consider that item difficulty affects item discrimination in that items of high and low difficulty may have lower discrimination statistics.
→ Compared with other incorrect alternatives, few examinees selected incorrect alternative D for this item. Consider reviewing the content of this alternative to determine if it can be revised to attract more examinees.

The first summary statement for this item states that the item is of moderate difficulty. If we look up to the information bar, the difficulty for the item is listed at 0.598, which means that 59.8% of students who took the test selected the correct (key) response (in this case “B”) to this item.

The second statement has to do with the discrimination of the item. In general terms, item discrimination refers to how well items differentiate (i.e., discriminate) between students of different ability levels. The higher the discrimination statistics, the better the item differentiates between the top-, middle-, and low-performing students in your class. For more information on the concept of item discrimination, click on the term within Integrity. The summary statement offers some direction in terms of what to look for in your item to try to evaluate and improve the performance of the item.

The third summary statement for this example relates to the performance of one of the item alternatives: option D. The summary statement describes a potential problem with the item in that very few students selected option D compared with the other incorrect alternatives (A and C). Upon reviewing this alternative, you may find that it is too obviously not the correct answer and may wish to edit it to make it more appealing to students.

By reviewing the content of the item-specific summary statements, you can gain a better understanding of the strengths and weaknesses of items and how to improve items.

It is highly recommended that clients review the more detailed guides that Integrity offers in the **Knowledge base** section in order to acquire more information on methods to evaluate and improve the quality of assessments.

## Other useful information produced by Integrity

There are many other useful features that Integrity offers, a few of which are described below.

### *PDF versions of reports*

Integrity produces PDF version of all web based reports at the discretion of the user (you). If you would like a “print friendly” version of the PDF reports simply select this option when submitting jobs to Integrity. You can then access the PDF version on the download page, which is labeled as a link on the left column of the application. The PDF versions of the reports are useful as an archive which you can save onto your local drive and review at your discretion. Once you have created and saved a version of the PDF reports you can delete the job online to prevent your Integrity homepage becoming cluttered with many tests that you have submitted to the system.

### *Data files*

Integrity produces data files which are comma separated files that contain all statistical information produced by Integrity. The production of data files is useful as an archive of all the statistical results produced by Integrity. You can read the data files into spreadsheets or for more sophisticated users the information can be read into analysis tools or item bank software.