

Quick start guide for using subscale reports produced by Integrity

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

Integrity refers to groups of items that measure a similar contextual area as “subscales.” For example, on a math test, you may have questions assessing three different areas of the curriculum. Items 1 through 25 may test the curricular area “Patterns and relations,” items 26 through 70 “Shape and space,” and items 71 to 100 “Statistics and probability.” Integrity lets you specify these three areas as “subscales” so that you can receive specific information on each area. If you would like to specify subscale information when submitting your job you should click on the “Would you like to specify subscale information?” option and follow the instructions on how to do so.

An example of a main subscale report page is shown in Figure 1. This page lists the results for each subscale, specified by you when submitting your test to Integrity. The table shows the label of each subscale (specified by you) the number of examinees, mean, standard deviation (SD), standard error of the mean (SE mean), standard error of measurement (SE measurement), and KR-20 for each subscale. In the example, we see that students performed best on the “Shape and space” items, followed by the “Statistics and probability” items, followed by the “Patterns and relations” items. By clicking on any of the subscale titles (e.g., Statistics and probability), you will be taken to a specific report that summarizes the performance of each subscale.

Figure 1. An example of the main subscale report page produced by Integrity.

Subscale	Number of examinees	Mean	SD	SE mean	SE measurement	KR-20
Patterns and relations	512	15.053	3.400	0.150	2.248	0.563
Shape and space	512	27.508	6.023	0.266	3.065	0.741
Statistics and probability	512	18.482	4.025	0.178	2.526	0.606

The specific report for the subscale “Statistics and probability” subscale is shown in Figure 2 below. The report has four sections:

- 1) summary statements
- 2) Table of Statistics for Subscale
- 3) Item Level Statistics for Subscale
- 4) Frequency Histogram for Subscale

The summary statements, as in other reports of Integrity, summarize the performance of the subscale and direct you investigate specific issues related to the subscale. In this

example, the first statement directs you to investigate the KR-20 reliability of the subscale. As the first summary statement indicates one factor that affects reliability is the number of items. In this case, the “Statistics and probability” subscale has only 30 items, many fewer than the 100 items that compose the entire test. This may be one factor contributing to the low subscale reliability. If it is determined that this is a contributing factor, consider adding more items to the subscale in a future version of the test. Another factor may be that the items on the subscale are not measuring one dominant trait or construct. It would be useful to review the content of each item in the subscale to ensure that items are in fact “Statistics and probability” items and are not measuring multiple curricular areas. The Phi correlation coefficients, provided in the item statistics summary report, can shed light on the statistical relationship of items to one another. The Pearson correlation coefficients and scatter plot section, discussed later, provide other methods of evaluating how related subscales are to one another.

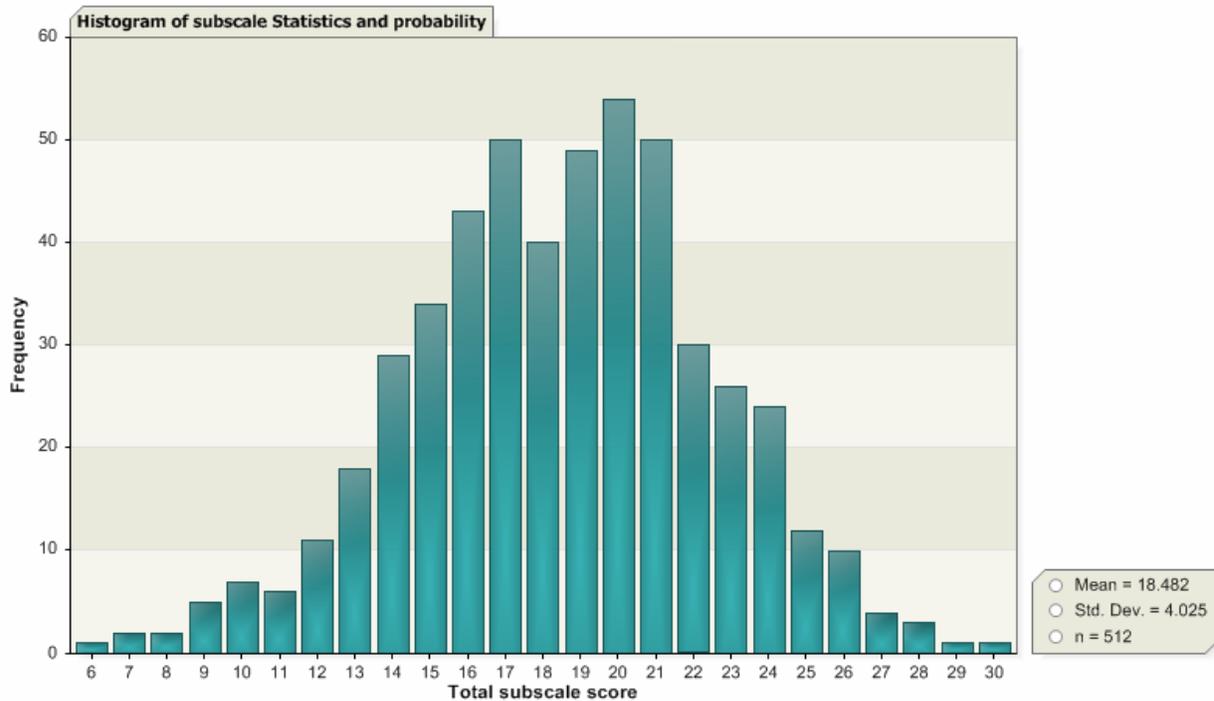
The second summary statement simply states that the distribution of subscales scores is relatively normal. The Histogram of Subscale graph verifies this statement visually.

The Item Level Statistics for Subscale table provides item level information specific to the subscale. For example, the discrimination (CPBR) of items will be different in this table than in the assessment level item summary table. This is because the discrimination statistic in the subscale report uses the total *subscale* score rather than the total test score in the calculation. In other words, the discrimination statistic for item 71 in the example below is correlating students’ subscale total scores (i.e., the sum of the correct items for the 30 items that compose the subscale) to their correct/incorrect performance on item 71. If the discrimination statistics for an item is low in the subscale report, it suggests that the item has a weak relationship to what the rest of the items in the subscale are measuring. The “KR-20 if item deleted” also reflects this relationship: if this number is greater than the KR-20 for the entire subscale, then the item did not perform well in the subscale. By examining the performance of each item you can gain more information as to how the items that compose your subscales are performing in relation to the subscale as a whole.

Figure 2. An example of a subscale report for a specific subscale produced by Integrity.

Summary: Statistics and probability			
–The KR-20 for this test indicates low test reliability. The reliability of the test is related to factors such as: 1) low number of test items, 2) small number of examinees, 3) many items that are too difficult to too easy, 4) many items that have low discrimination, 5) the items on the test are not measuring one dominant trait, 6) students do not have enough time to finish all the items on the test. Consider investigating the above factors in order to increase test reliability.			
–The skewness and kurtosis statistics indicate that the distribution of scores is relatively normally distributed.			
Table of statistics for subscale: Statistics and probability			
Number of examinees = 512	Standard error of mean = 0.178		
Number of items on test = 30	Standard error of measurement = 2.526		
Mean = 18.482	KR-20 reliability = 0.606		
Median = 19.000	Spearman-Brown split-half reliability coefficient = 0.596		
Mode = 20.000	Spearman-Brown prophecy reliability formula = 0.747		
Standard deviation = 4.025	Guttman split-half reliability coefficient = 0.595		
Variance = 16.199	Skewness (total score) = -0.154		
Maximum score = 30	Kurtosis (total score) = 0.016		
Minimum score = 6			
Item level statistics for subscale: Statistics and probability			
Item number	Difficulty	Discrimination (subscale CPBR)	KR-20 if item deleted
71	0.584	0.199	0.578
72	0.678	0.267	0.573
73	0.480	0.193	0.578
74	0.844	0.298	0.579
75	0.643	0.251	0.574
76	0.662	0.231	0.576
77	0.623	0.166	0.583

Frequency histogram for subscale: Statistics and probability



An additional piece of information available in the subscale analysis section is the correlation results between subscales and corresponding scatter plots. An example of this information is shown in Figure 3. The purpose of this analysis is to obtain a sense of how much overlap there is in the measurement of subscale constructs (i.e., how related the scores from each subscale are to one another). If two subscales are highly related (and therefore highly correlated – for example 0.910), it may make more sense to collapse the two subscales into one category rather than interpret the results for each subscale separately. The scatter plots present a graphical representation of the correlation coefficient; the more the individual points are clustered together in a cigar-shaped manner, the higher the correlation coefficient. Positive correlation coefficients indicate that as one factor (e.g., scores for subscale “Patterns and relations”) increases, the other factor (e.g., scores for subscale “Shape and space”) increases as well. Negative correlation coefficients indicate that as one factor increases, the other decreases.

Figure 3. An example of the Pearson correlation coefficients and scatter plots for subscales produced by Integrity.

Subscale total score Pearson correlation coefficients			
→The correlation between subscale Patterns and relations and subscale Shape and space is in the large effect size range.			
→The correlation between subscale Patterns and relations and subscale Statistics and probability is in the large effect size range.			
→The correlation between subscale Shape and space and subscale Statistics and probability is in the large effect size range.			
	Patterns and relations	Shape and space	Statistics and probability
Patterns and relations	1		
Shape and space	0.653 (p = 0.000)	1	
Statistics and probability	0.58 (p = 0.000)	0.657 (p = 0.000)	1

Scatter Plot: 2	
-----------------	--

