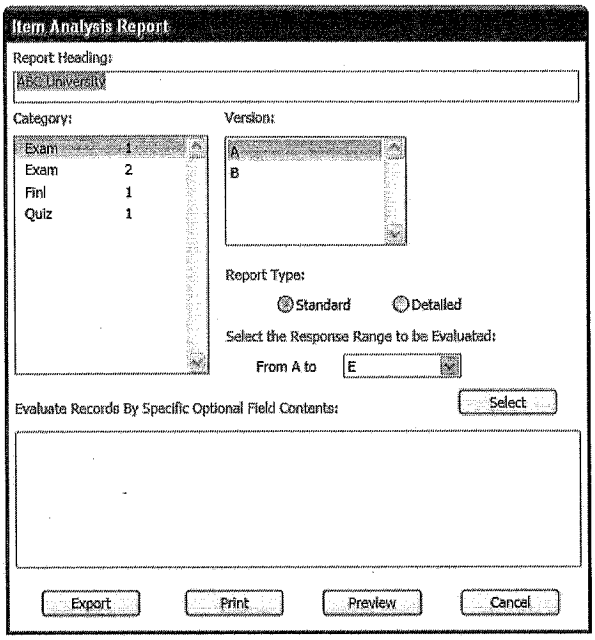
Item Analysis Report:

This is a statistical report that provides a detailed distractor analysis based on raw scores. Each test version is generated separately. See page 6-12 for details about Item Analysis data.

1. Select the Category.
2. Select Print or Preview.

Optional:

3. Select the Report Type (Standard or Detailed).
4. Change the Response Range if needed (choices beyond "E").



ABC University

Standard Item Analysis Report On Exam1 Version A

Course #: NRS155 Instructor: Dr. Alec Small
Course Title: Foundations of Nrsg Description Foundations of Nursing
Day/Time: Term/Year:

Total Possible Points:	10.00	Median Score:	8.50	Highest Score:	10.00
Standard Deviation:	1.34	Mean Score:	8.17	Lowest Score:	6.00
Student in this group:	6	Reliability Coefficient (KR20): 0.39			
Student Records Based On:	All Students				

No.	Correct Group Responses			Point Biserial	Correct Answer	Response Frequencies - * indicates correct answer										Non Distractor
	Total	Upper 27%	Lower 27%			A	B	C	D	E						
1	83.33%	100.00%	50.00%	0.72	C	1	0	*5	0	0						BDE
2	83.33%	100.00%	50.00%	0.72	A	*5	1	0	0	0						CDE
3	100.00%	100.00%	100.00%	0.00	A	*6	0	0	0	0						BCDE
4	83.33%	100.00%	50.00%	0.72	C	1	0	*5	0	0						BDE
5	83.33%	50.00%	100.00%	-0.28	A	*5	0	1	0	0						BDE
6	66.67%	100.00%	50.00%	0.61	D	1	1	0	*4	0						CE
7	83.33%	100.00%	100.00%	0.06	D	0	0	1	*5	0						ABE
8	66.67%	100.00%	50.00%	0.09	D	0	0	2	*4	0						ABE
9	83.33%	100.00%	50.00%	0.39	A	*5	0	1	0	0						BDE
10	83.33%	100.00%	50.00%	0.39	A	*5	0	0	1	0						BCE

Taking a look at the Item Analysis data:

After an objective test has been administered and scored, it is desirable to evaluate the effectiveness of the items. In order to improve items, it is necessary to examine whether or not they are doing the job for which they were designed. An item analysis provides four kinds of important information about the quality of the test items:

Reliability Coefficient (KR_{20}): The *Kuder-Richardson* formula 20 (KR_{20}), for example, calculates a reliability coefficient based on the number of test items (k), the proportion of the responses to an item that is correct (p), the proportion of responses that are incorrect (q), and the variance (σ^2). Consistency is a measure of reliability through similarity within the test, with individual questions giving predictable answers every time.

The *Kuder-Richardson reliability* assesses inter-item consistency of a test by looking at two error measures:

- Adequacy of content sampling
- Heterogeneity of domain being sampled

It assumes reliable tests contain more variance and are thus more discriminating. Higher heterogeneity leads to lower inter-item consistency. The Reliability Coefficient should fall between 0.4 and 0.7 for best results.

Item difficulty is the percentage of the total group that got the item correct. The item difficulty is important because it tells you whether an item is too easy or too hard. Many test experts believe that for a maximum discrimination between high and low achievers, the optimal level is 50%. However, because of the guessing factor, it is advisable to make a test somewhat easier. This reduces the guessing factor and consequently increases the test reliability. For true-false items a 75% difficulty level is recommended. That is, "on the average," 75% of the students should get the item correct.

Multiple-choice items:
3 alternative multiple-choice = 67%
4 alternative multiple-choice = 63%
5 alternative multiple-choice = 60%

Items that have percentages less than 30% or more than 90% definitely need attention. Items that have difficulty levels that are too hard or too easy should either be revised or replaced. The only exception to this standard occurs with items that appear at the beginning of a test. The first few items should be easy (Difficulty level = 90% or higher) for psychological reasons.

To identify items as easy, medium or difficult, the following helps to read the Item Analysis Report. The first column refers to the difficulty:

1. If more than 75% of the students get the question correct, the item is considered easy.
2. If 50 to 74% of the students get the item correct, the item is considered medium in difficulty.
3. If less than 50% of the students get the item correct, the item is considered difficult.

Item Discrimination is the single best measure of the effectiveness of an item is its ability to discriminate between students who vary in their degree of knowledge of the material tested. If we had two groups of students, one of which was composed of students who had mastered the material and the other of students who had not, we would expect a larger portion of the former group to correctly answer any test item. Item discrimination attempts to give us this information by measuring how well an item discriminates between these two groups.

ParScore provides the percentage of the upper and lower groups that got a particular item correct.

If desired, you can calculate the item discrimination ration by using the following simple formula:

IDR = (Upper Group % Correct) - (Lower Group % Correct)

The maximum item discrimination ration of an item occurs if the discrimination ration = 100%. This would occur if all those in the upper group got the item right and none of those in the lower group got it correct. Zero discrimination occurs when equal numbers in both groups got it correct. Negative discrimination occurs when more students in the lower group than the upper group get an item correct. Zero and negative discrimination items should be discarded or vastly improved before the item is used again on the next test.

An acceptable level of discrimination for classroom tests is at or above 25%. An item having a ratio lower than this would be considered a poorly discriminating item. That is, it would not discriminate between those who really knew the material and those who were less knowledgeable. Items having ratios above 40% are considered excellent items.

Columns two and three (Upper and Lower 27%) refer to the discrimination. The easiest way to calculate the discrimination at a glance when viewing the Item Analysis Report is the following:

Upper 27% - Lower 27% = Discrimination

When looking at the differences the following holds true:

1. If the discrimination is 40% and above, the discrimination is high. This is preferred.
2. If the discrimination is 25% to 39%, the discrimination is medium. This is acceptable.
3. If the discrimination is below 25%, the discrimination is low. This indicates the item should be reworked or it is material you expect your students to know.

Point Biserial Correlation Coefficient (PBCC) measures the correlation between the correct answer on an item and the total test score of a student. The PBCC is the preferred method of measuring item discrimination because it identifies items that correctly discriminate between high and low groups, as defined by the test as a whole (instead of only the upper and lower 27% of a group). The following criteria may be used to evaluate test items:

1. .30 and above — Very good item
2. .29 to .20 — Reasonably good item (subject to improvement).
3. .19 to .09 — Marginal items (needs improvement).
4. below .09 — Poor items (reject or improve).

Rescoring a Test: A Quick Fix: ParScore provides a powerful tool for resolving (or at least partially resolving) problems. After studying the item analysis report, the answer key may need to be modified. Items can be removed from the test or automatically given to all students. Inspection of distracters would suggest that perhaps the wrong answer was marked on the key. This item may need to have the answer changed or possibly more than one answer should be allowed to be correct.

Review the columns labeled "Response Frequencies." This provides information on the number of students that chose each alternative. If needed, return to the answer key tab and modify the desired items. The test will be rescored automatically, without having to rescan the student's test forms.

Distracters and their effectiveness: By looking at the pattern of responses to alternatives, the instructor can often determine how the test can be improved. A multiple-choice question is only as good as its distracters. If two distracters in a four-choice item are completely implausible, the question is, in effect, a two-choice or true-false item and made easier. It is important for an instructor to know if the distracters are really "distracting" or just taking up space on paper. ParScore automatically identifies those distracters of each item that did not "distract" students.