

# ExamSoft Item/Exam Analysis Cheat Sheet

## Analyzing Distractors/Choice Frequency

The general item analysis will assist you in determining whether a distractor is good. As you can see below for item 1, A is a poor distractor since no one selected that answer. For example in item 3; 2 of 4 distractors are implausible, so it is easy to answer. It is important a distractor be "distracting."

Item	A	B	C	D
1	0	4	15*	4
2	10	14*	3	1
3	2	23*	0	0

\*Correct Answer

## Difficulty Index (P-value)

Simply divide the number who got it correct by the total number of students. A good "rule-of-thumb" is that if the item difficulty is more than .75 (75%), it is an easy item; if the difficulty is below .25 (25%), it is a difficult item.

### SAMPLE EXAM SCORES

Student	Score	Q1	Q2	Q3
1	90	■	-	■
2	90	■	-	■
3	80	-	-	■
4	80	■	-	■
5	70	■	-	■

Item	# Correct Upper	# Correct Lower	Difficulty Index (p)
1	4	4	.80
2	0	3	.30
3	5	1	.60

## Discrimination Index

Measures the item quality when you want aitem that is not too easy or too hard. It is the top percentage of the highest scoring students in the overall exam that scored correctly divided by the lowest percentage. The closer to 1 you get, the better the item is.

Item	# Correct Upper	# Correct Lower	Discrimination Index
1	4	4	0

2	0	3	-0.6
3	5	1	0.8

Item #2 had a difficulty index of .30 (quite difficult), and a negative discrimination index of -0.6 (low-performing students likely to get item correct). This item should be carefully analyzed, and probably deleted or changed. The "best" overall item is Item 3, which had a moderate difficulty level (.60), and discriminated extremely well (0.8).

## Point Biserial

Indicates that exam takers who performed well on the exam also selected the correct response, so this is a good discriminator between high-scoring and low-scoring students.

Item	Upper27%	Lower27%	Point Biserial
5	66.7%	33.3%	.42

### General Interpretation

Very Good Item: .30 and above  
Reasonably Good: .20 - .29  
Marginal Item: .09 - .19  
Poor Item: below .09

## KR-20 (Kuder-Richardson Formula 20)

This statistic measures exam reliability of inter-item consistency. A higher value for the exam indicates a strong relationship between items on the test. This shows the reliability (indication) that the same students taking the test again, would achieve the same scores. A value of at least .70 is desirable. Range: 0.00 – 1.00.

## Standard Deviation (stdev)

This is a measure of how spread out the distribution of scores is on an exam. Range: 0.00 – 1.00. A low stdev indicates that most of the testers are close to average. A high stdev indicates that there is a wide variation of scores within the exam. In order to consider an exam good or bad, you need to define limits. What % of the students fall within the limits and how many failed to make it in the limits. If the failure rate is higher then the limit, then you would rate the exam "not good".