EDUCATIONAL DEVELOPMENT CENTRE

Scantron Guides

Item Analysis

Item analysis refers to a statistical technique that helps instructors identify the effectiveness of their test items. In developing quality assessment and specifically effective multiple-choice test items, item analysis plays an important role in contributing to the fairness of the test along with identifying content areas that maybe problematic for students.

Generally, the process of item analysis works best when class sizes exceed 50 students. In such cases, item analysis can help in identifying potential mistakes in scoring, ambiguous items, and alternatives (distractors) that don't work. When performing item analysis, we are analyzing the following important statistical information:

Proportion answering correctly (item difficulty) indicates the proportion of students who got the item right. A high percentage indicates an easy item/question and a low percentage indicates a difficult item. In general, items should have values of difficulty no less than 20% correct and no greater than 80%. Very difficult or very easy items contribute little to the discriminating power of a test.

Question 1

Prop. Answering Correctly: 0.50
Discrimination Index: 0.57
Point Biserial: 0.43

Alt.	Total Ratio	Low (27%)	High (27%)	Point Biserial	Key
A	0.14	0.27	0.10	-0.28	
B	0.25	0.30	0.07	-0.15	
C	0.50	0.23	0.80	0.43	*
D	0.10	0.20	0.03	-0.18	

Discrimination index is the difference between the proportion of the top scorers who got an item correct and the proportion of the bottom scorers who got the item right (each of these groups consists of twenty-seven percent of the total group of students who took the test and is based on the students' total score for the test). The discrimination index range is between -1 and +1. The closer the index is to +1, the more effectively the item distinguishes between the two groups of students. Sometimes an item will discriminate negatively. Such an item should be revised and eliminated from scoring as it indicates that the lower performing students actually selected the key or correct response more frequently than the top performers.

Continued on reverse.

Point biserial is the correlation between an individual student's performance on an item and his or her total score on the test. The values range from -1 to +1. The high positive values are desirable for the correct answer because they indicate that a student who did well on the exam also did well on this question. Negative values are desirable for the alternatives or distractors that were not the correct answer. A score of 0 or less, for the correct alternative, indicates the question has difficulty distinguishing between those students who know the material and those who do not. The question should be examined and revised and potentially eliminated from scoring.

Distractor evaluation is another useful step in reviewing the effectiveness of a test item. All of the incorrect options, or distractors, should actually be distracting. Preferably, each distracter should be selected by a greater proportion of the lower scorers than of the top group. In order for a distractor to be acceptable it should attract at least one candidate. If no one selects a distractor it is important to revise the option and attempt to make the distractor a more plausible choice.

Reliability of the test - the summary statistics found at the beginning of your item analysis includes an estimate of the test's reliability. KR 21 is a measure of the internal consistency of your exam or how well the items work together to obtain a measure of student achievement. Generally, values of .60 or greater are acceptable for the purposes of classroom tests; however, an accurate interpretation of this value frequently requires the consideration of a number of factors impacting on student performance. Factors which lower the reliability of a test include: items which are poorly written, too many items which are very easy or very hard and too few items overall on a test to obtain a stable estimate of the student's ability.

ITEM ANALYSIS

Professor: EDC
Course Code: EDC
Date of Exam: 01/01/09

# of Questions: 50
# of Students: 111
# of Keys: 1
Weighting: 1.0

Mean: 32.78
Variance: 48.13
Std. Deviation: 6.94
Skew: -0.38
Kurtosis: -0.16
Lowest Score: 12.00
Highest Score: 46.00
Median: 33.00
Mode: 33
Reliability: 0.78
Std. Error of Measurement: 3.25

Bonus Questions:
Questions Eliminated: 21, 24
Exam Marked Out of: 48

(Adapted from Michigan State University website & Barbara Gross Davis *Tools for Teaching*)

Educational Development Center
410 Dunton Tower
edc.carleton.ca

Phone: 520-4433
Fax: 520-4456
E-mail: edc@carleton.ca