

Activity 5: Brief Review of Test Item Analysis

The material below is a brief review of item analysis.

I. Introduction

A. Item analysis is a series of techniques used to

1. Improve tests and test items
 - a. revising existing questions is more efficient than creating new ones
 - b. previously developed items usually provide a better measure of student achievement than untried, unrevised items
2. Increase the validity of tests
3. Provide data for instructional decision-making

B. These techniques are most commonly used with objectively scored items, especially multiple-choice items

II. Three Components of Item Analysis

A. Item Difficulty

1. This is a measure of the difficulty of the item and it is measured by calculating the proportion (percent) of students answering an item CORRECTLY. The higher the proportion answering correctly the LESS difficult the item is (this is somewhat like the higher your score on a test, the less difficult the test is)
2. The value ranges from 0 – 100% (0 – 1.00)
3. The value is calculated by dividing the number of students who answered an item correctly by the total number of students taking an exam.
4. Example: If 20 students take Random Exam A and 14 students answer item 5 correctly, then the item difficulty is $14/20$ or .70 (70%). On this same test, 3 students answer item 10 correctly, so the difficulty for item 10 is $3/20 = .15$ or 15%. Thus, Item 5 is an easier item than item 10.

B. Item Discrimination

1. This is a measure of how well an item distinguishes between those students who know the content assessed in a given item and those who do not know the content.
2. The value of this indicator ranges from 0 – 100
3. The value is calculated by using the following procedure:
 - a. Score all tests.
 - b. Identify two subgroups of test takers, the upper and lower 1/4, or the upper and lower 1/3, or the upper and lower 1/2. Note, for classroom assessment, the tests are often divided into upper and lower 1/2 or upper and lower 1/3 given the small sample size.
 - c. For each test item, determine the proportion of students in the upper 1/2 (or whatever fraction is used) of the group (based on the overall test scores) answering the item correctly [difficulty index for the upper 1/2], and the proportion of students in the lower 1/2 of the group answering the item correctly [difficulty index for the lower 1/2].
 - d. Subtract the lower proportion from the upper proportion. A discrimination index of 20 or 20% or .20 indicates that the item is doing an adequate job of differentiating between the upper and lower half of students. NOTE: Item discrimination is used more widely for identifying items to place on standardized tests than on classroom tests. In your classroom, you may want to calculate item discrimination simply to have a better sense of who is struggling with an item – all students equally (discrimination value of less than 20%), the lower group (discrimination value of 20% or higher) or the lower group (negative discrimination value)
4. Example: Item 15 on Random Exam A. 20 students took this test. The difficulty index for the upper 1/2 of the group is 80% (which means 80% of students answered this item correctly), and 40% of the lower half answered item 15 correctly. This provides a discrimination index of $80\% - 40\% = 40\%$ for item 15 and indicates that this item was missed more by those students who generally did not do as well on the test as a whole.

5. Example: Item 16 on Random Exam A. 20 students, difficulty for upper $\frac{1}{2}$ is 30%; for lower half, it's 90%. Discrimination index is $30\% - 90\% = -60\%$. This means that more students in the upper $\frac{1}{2}$ of the scorers missed this item. This is not typical, and MAY (but not always) indicate a problem with the test item rather than with the understanding of the content itself.

C. Distractor Analysis

1. This involves looking at the options for a given item and examining the percentage of students in each group choosing each option.
2. When an item has GOOD distractors:
 - a. The distractors distract more frequently the students in the lower $\frac{1}{4} - \frac{1}{2}$ of test scores, than those in the upper $\frac{1}{4} - \frac{1}{2}$.
 - b. All distractors are reasonable choices for a student who DOES NOT know the content, but are not so close that those who do know the content are confused by them.
 - c. When possible, use distractors from the same category as the correct response. For example, if you are asking students to identify the part of speech for the underlined word in a sentence, and it's a noun, then good distractors might include pronoun, verb, and adjective. If you are asking students to classify an animal into a subgroup of vertebrates, and the animal is a reptile, then plausible distractors are mammal, bird, and amphibian.
3. Careful distractor analysis can help a teacher identify misconceptions or misunderstandings about content. For example, in the parts of speech example, if the correct response is NOUN and no students select verb or adjective, but 5/20 select PRONOUN, then you know that students understand that the part of speech represents a person, place or thing, but that they are confused between the actual name of the noun and the substitute or stand in for the noun. That is, they are treating BOB and HIM as the same part of speech. This would lead you to reteach on the difference between a noun and a pronoun (which can stand in for a noun).

***It's important to keep in mind that with 50 or fewer students, as with almost all classroom assessments, the values for all 3 components of the item analysis are unstable, so you look at ball park figures and make instructional decisions that make common sense and are not based on a rigid interpretation of numerical data.

III. **Interpreting an Item Analysis**

1. In what order does one look at the three components? Look at difficulty, then discrimination (maybe – this is the least important component for classroom assessment and tests that are created by the teacher) followed by distractor analysis.
2. If more than $\frac{1}{2}$ of your class misses an item, it is likely that you need to reteach the content, particularly if the content being assessed is a very important concept. If less than $\frac{1}{2}$ the class misses an item, you still may need to provide differentiated reteaching for those students who missed the item.
3. If, upon careful inspection of the test item, it appears that the item itself is faulty, then you would not need to reteach the content, but you would want to revise the item before using it in the future AND I recommend that you adjust student grades in accord with the problems you found with the item.
 - a. Example: A multiple choice test item with one correct response and three distractors is missed by about $\frac{3}{4}$ of the class. As you look carefully at the item, you realize that NONE of the answers is actually correct. Those students who answered “correctly” chose B, the answer you marked as correct. However, B is also not correct. In this case, you would NOT count that item against any students and you would drop it from your test. So, on a 20 item test, you now have only 19 items and a student's score is calculated by dividing the number answered correctly (raw score) by 19 instead of 20.
**If a student chose C for this item and they answered 17 items correct, then their score is calculated as 17 of 19 (89%) instead of 17 of 20 (85%). The number they answered correctly has not changed, but the possible number of items on the test has changed.

**If a student answered B (indicated as correct on your key, but not really correct) then the student who had 17 of 20 (85%) correct now has 16 of 19 (84%) correct (one less item, but also one less item marked as correct).

4. Using Student Input to Interpret Item Analysis

1. why this is valuable:
 - a) students read the test carefully and often detect more ambiguities than the teacher can
 - b) also, reading items from perspective of learner rather than that of teacher
2. student comments can lead you to see what misconceptions they had about item, so discuss problematic items with students after viewing the item analysis
3. can lend insight into problem items as well as instructional problems
4. generally, discuss only the most difficult items