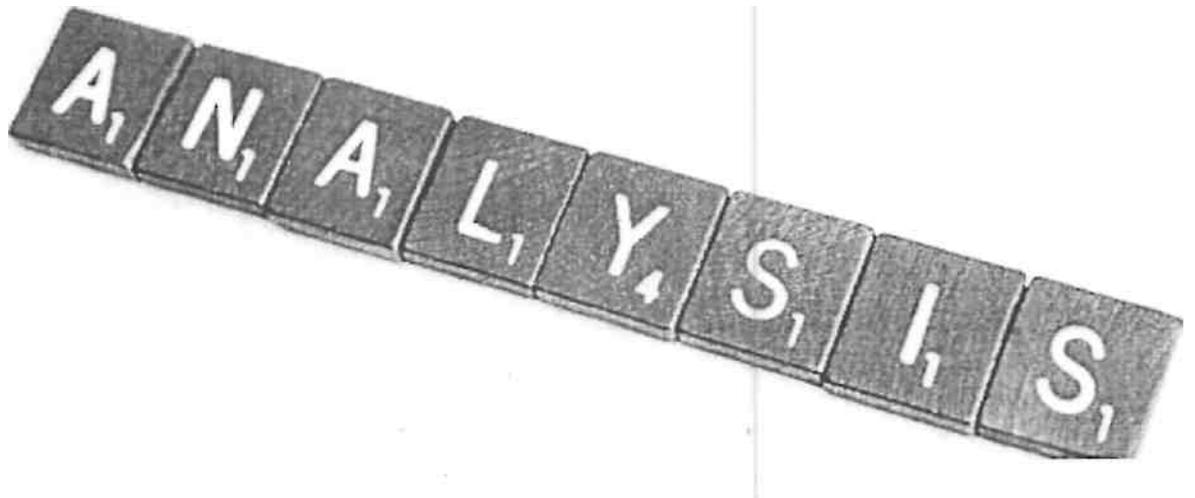# Improving Multiple Choice Questions Through Item Analysis

September 29, 2014 / 0 Comments / in Blog /
by Michelle Watts

Analyzing an item typically requires a deeper look at three components: item difficulty, item discrimination and response distribution. The item difficulty, also called the p value, is the proportion of students who correctly answered an item, while the item discrimination is the ability of the item to separate the students who did well on the test from those who did poorly on the test. Lastly, the response distribution is the percent of students who selected each answer choice. These statistics can provide valuable information for test item revisions and enable curriculum and instructional improvement.

Frequently, the p value is the only item statistic that educators examine when reviewing test item statistics. While the p value does provide useful information, the information provided by the aforementioned three components can provide educators with a more comprehensive analysis.   Most educators create and use criterion-referenced tests that are designed to measure what students knows when compared with a standard such as content mastery.   With that purpose in mind, the p

value for most items on the test should fall between .60 and .90. While a test might contain a small number of items that have a higher or lower p value, that number should be small. When a test item has a p value less than .60, which means that less than 60% of test takers correctly answered the item, the other components of item analysis can provide useful information to diagnose problems with the item.

Let's examine the item statistics for a test item that is appropriate for a criterion-referenced test. Item 25 on the test was correctly answered by 82% of the students. That p value, .82, is within the recommended range of .60 to 90 for a criterion-referenced test. The answer choice distribution indicates that students selected each of the answer choices. None of the distractors (incorrect answers) was favored or selected more frequently than the other distractors.

Item	Answer A	Answer B	Answer C	Answer D	p	Discrimination
25	82%	4%	6%	8%	.82	.26

The item discrimination value can range from -1.00 to +1.00. Generally a negative discrimination indicates a problematic item. When more students who did poorly overall on the test correctly answer the item more frequently than students who earned a high score on the test, the item discrimination value will be negative. A positive item discrimination means that more students with high test scores correctly answered the item than student with low test scores. The higher the discrimination value, the better an item is at separating the high performing students from the low performing ones. A discrimination value greater than .15 indicates a reasonable rate for a criterion-referenced test.

Now let's review some problematic items. Test item 17 has a p value of .52 and answer choice A is the correct answer. This item is slightly difficult for a criterion-referenced test. The item has a good discrimination value, but a review of the answer choice distribution indicates that answer choice B was a favorite error made by students. To improve this

item, review answer choice B to determine if the answer is partially correct or too close to the correct answer for the students to identify the difference.

Item	Answer A	Answer B	Answer C	Answer D	p	Discrimination
17	52%	31%	10%	7%	.52	.27

If answer choice B on item 17 is a common error that students make, the teacher now has specific information on student errors. Rather than revising the item, the teacher can now modify and plan instruction accordingly to prevent that error in the future. The teacher can re-teach the content along with test-taking strategies. If the item is not a secure item, the teacher can use the test item as a way to model the metacognitive procedures students can use to eliminate answer B. Ask the students to explain why each distractor is wrong and why the keyed answer is correct.

Test item 8 has a different problem. The item is very difficult for the students as only 36% of the students selected the correct answer as B. This item falls outside the recommended range for the item difficulty of .60 to .90 and has a discrimination value below the .15 suggested minimum value for a test item. An examination of the distractor distribution indicates that the students selected all of the incorrect answer choices at a similar rate. Although the item is difficult, what else is wrong with the item? When students select all the answer choices at a similar rate, including the correct answer, the item is unclear to the students or the information has not been taught. They are basically guessing when answering the item. With this item, the stem or the question is the likely problem. Review the stem for clarity and sufficient information. The question should be revised or additional information should be added.   If the stem is clear, it is possible that the students have not received instruction on the content measured in the item.   If that is the case, the teacher has clear information on what the students need to be taught or re-taught.

Item	Answer A	Answer B	Answer C	Answer D	p	Discrimination
8	27%	36%	22%	15%	.36	.067

At first glance item 4 looks like a good item. The p value is .68, within the recommended range, and the students selected all of the answer choices. The discrimination index is -.246 which is greater than the recommended value of .15. So what is the problem with the item? The discrimination value is negative, which means that more of the students who received a low score on the test answered the item correctly than did students with a high score on the test. The problem is likely answer choice B. Most likely the answer choice is plausible or makes sense to students with a high degree of knowledge. Answer choices that are plausible are frequently found by students who have a high rate of achievement. They are more likely to think about the item differently than would average achieving students. To improve this item, answer choice B should be revised. If possible, ask the students why they selected the distractor. Their reasoning can provide the test user with interesting information about the thinking process of the students.

Item	Answer A	Answer B	Answer C	Answer D	p	Discrimination
4	8%	24%	68%	4%	.68	-.246

Item analysis data can provide educators and administrators with valuable information to improve item quality and, ultimately instruction. The overall test score is a useful indicator of student achievement, but the picture is more complete when item analysis is also examined.

MICHELLE WATTS [HTTP://WWW.THINKGATE.COM/AUTI

Michelle Watts holds a Ph.D. in Measurement and Evaluation from the University of South Florida in Tampa, Florida. Dr. Watts is currently an education consultant and adjunct instructor, teaching Statistics at Embry Riddle Aeronautical University. Recently retired from Hillsborough County Public Schools, Dr. Watts worked for 25 years in the Assessment and Accountability Department supervising the district's test development activities. The department was responsible for the creation and delivery of over 700 examinations each semester to the district 175,000 students.