

Module 4C: Analyzing Item Difficulty and Discrimination

by Dr. Tim Brophy

Hello. My name is Tim Brophy. And welcome back to Passport to Great Teaching-- Creative Assessment. This is module 4C, the third and final video of module 4. And in this module, we're really going to dig into analyzing item difficulty and discrimination.

So we know that teachers who create tests for classroom use, like most of us do, often seek to know how effective their tests or quizzes are. So item analysis gives us a way to really know how effective our tests and quizzes are by providing some important information about how well the items actually function. So item difficulty helps us to know the degree to which students get answers correct, whereas item discrimination examines how the top-scoring group of test takers compares to the lowest-scoring group of test takers, another important piece of information to help us know how well our items are working.

So let's talk about calculating item difficulty. Now, to do this, what we first have to do is count the total number of students who answer the item correctly-- not the test, but the item itself on the test. Then for every item, we're going to divide the number that answers correctly by the total number of students. This is going to give us the proportion of students who actually got the item correct, who answered it correctly. So this figure is called the "items difficulty level."

Now, you have to be careful when you're interpreting item difficulties because the way we calculate this is that the higher the difficulty level, the easier the item is and vice versa. So for example, if you have an item difficulty of 0.95, it doesn't mean it's extremely difficult. It means it's extremely easy because that 0.95 represents the fact that 95% of the students got it correct.

Now let's talk about item Discrimination Index, or DI. When we subtract the proportion of low-scoring students who got an item right from the proportion of high-scoring students who got it right, then the remainder becomes the discrimination index. Now, this is a measure of how well the item discriminates between the top scores and the bottom scores on the item.

So when we calculate the DI, the first thing we do is we have to compute each student's score. That is, we have to count the number of items each student got correct, which we do anyway. And then we divide the class into groups on the basis of their total scores.

For example, we could divide it into 50% high, 50% low. Or if there are a lot of students, we could have 33% high, 33% in the middle, 33% in the low group. Or you could pick 34 for

any one of those numbers. Usually it's 33, 34, 33, or 25% in the high group, 50% in the middle, and 25% in the low group, whatever way you think is best for your particular situation to help you determine the degree to which the top scores and the lowest scores on an item actually are different.

So the high and low groups have to be the same size, though. That is one of the requirements for this calculation. So the middle group really isn't needed in computing discrimination indices. So if you do divide into three groups, that middle group we don't really need because they're not part of this calculation.

Then in the fourth step, for every item, you have to count the number of students in the high-scoring group that got the item correct. And then in the next step, you divide that number about the total number in the high-scoring group for each item. Then you're going to repeat steps four and five for each item, this time only using the low-scoring group.

Then for each item, subtract the proportion of low-scoring students who got the item right from the proportion who got it correct in the high-scoring group. And that remainder becomes the Discrimination Index, or DI. So those are the steps that we go through to do this. And we're going to have some examples here in just a moment.

So how do we use difficulty and item discrimination indices for item analysis? Well, here's what we suggest. Difficulty really ought to range between about 0.25 to 0.75 for most items.

Why? Well, items that are really below the 0.25 threshold are really the most difficult. And they may not be doing a good job of being an effective item for you. Items above 0.75 really cross the probability threshold for guessing, especially if you have four answer choices on your test.

Now, the lowest recommended discrimination index is 0.20. Why? Well, if there's less than a 20% difference between students in the top-scoring and low-scoring groups that respond to the item correctly, the item is really not discriminating between those high and low scores as effectively as you'd want it to do.

So those are just general guidelines. So they're not hard and fast rules. So it really is up to you to make the decision as to what you think is the best difficulty level and discrimination index for the items that you have on your test once you do these analyses.

OK. So now let's work an example. Here are the results of part of a biology test that was given to 100 students. So this was a 13-item test. But we're going to take a hard look at the first three items and determine their difficulty level and discrimination indices. So let's begin with that.

So here you see we divided up the total number of scores into three groups-- the top-scoring group of 25 students, the middle-scoring group of 50 students, and the lowest-scoring group of 25 students. So here are the results.

So in the top-scoring group, 20 got it correct. Or that's 80% of the top-scoring group. And in the middle group, 20 got it correct. Now, we're not calculating the percentage who got it correct because we don't need that for the discrimination index. We do need it for the difficulty level. And then in the lower group of 25, 15 got it correct, or 60%.

So let's do a little calculating. In order for us to calculate a difficulty level, remember, all we need to do is add the total of students who got it correct. So that's 20 in the top-scoring group, 20 in the middle-scoring group, and 15 in the lower-scoring group, or 55. We divide that number by the number of students, and we get 0.55. So the difficulty level is 0.55.

And to get the discrimination index, we subtract that percentage of students in the top-scoring group who got it correct, or 80%. We subtract 60% from that. That's the percentage of students who got the item correct in the low-scoring group. And we get a discrimination index of 0.20.

We do the same thing in item number two. Here, the top-scoring group, 15 got it correct, or 60% of the students. In the middle-scoring group, 30 got it correct. But we're not going to calculate the percentage who got it correct there. And then in the lowest-scoring group, 5 got it correct, or 20%.

So let's calculate the difficulty level. We simply add the total who got the item correct. In the top-scoring group, that's 15. In the middle-scoring group, that's 30. And in the low-scoring group, that's 5. And that adds up to 50. And divide that by the number of students, or 100 students, and we get 0.50 as our difficulty level.

The discrimination index-- we simply subtract 20% from 60%. And we get 40%, for a discrimination index of 0.40.

Moving on to item number three, in this one, the top-scoring group, 24 got a correct, or 96% of the students. In the middle-scoring group, 38 got it correct. And we won't calculate that percentage for this calculation because we only need that for the difficulty level. And in the lowest-scoring group of 25, 12 got a correct, or 48%.

So to calculate the difficulty level, we simply add the total who got it correct. And that's 24 plus 38 plus 12, or 74 students. Divide by 100. And we get a difficulty level of 0.74.

Then the discrimination index-- we subtract the 48% who got it correct in the low-scoring group from the 96% who got it in the top-scoring group. We come up with a discrimination index of 0.48.

So now we have to think about, OK, we've got this information. Do we keep or do we revise the items? Well, let's talk a little bit about some guidelines for that. When an item meets your general acceptability guidelines for difficulty and discrimination, just keep the item. That's my recommendation. You don't need to change it.

But if it falls outside the guidelines, you've got to make a judgment whether or not you're going to keep it or revise it or reject it. So remember those guidelines that we gave you earlier in this module. So what are they? We'll review them again.

Remember, difficulty level-- the standard range is 0.25 to 0.75. That discrimination index-- the standard is usually 0.20 or higher. So when we look back at item one, we just calculated the difficulty as 0.55 and the discrimination index as 0.2. We should probably just keep that item. It's substantially difficult, and it discriminates OK.

Item two discriminates even better. It's a little bit more difficult. And then item three is not as difficult. But it still has a good discrimination index, so we'll definitely keep that. But remember, 0.74 is pushing the top end of that 0.75 range. But when we look at the discrimination index, when we see 48% more of the top-scoring group getting the item correct in the low-scoring group, then we want to keep that item as well.

So what if you get negative results? Well, your difficulty levels should never be negative. So if you get a negative result, you need to recalculate.

Now, item discriminations can be negative. And that simply means that more students in the lowest-scoring group got the item correct more often than the high-scoring group students. So that is really a sign you need to do something about that. So the general advice in that case is to reject, revise, or completely rewrite the items with negative item discriminations. You don't want your lowest-scoring group doing better than the high-scoring group on an item because that means something is probably incorrect or unclear on the item for your students.

A good thing for you-- if you're using Canvas, the nice thing is that Canvas calculates all of these for you. You don't have to do them by hand at all. And there is a link to the Canvas document about its item analysis features in this particular place where I've highlighted the URL.

So click on that, and you'll go to that. And you'll see what Canvas offers you. If you happen to have a test or an item or a quiz that you've given in Canvas, you can look up the item discrimination information and use that to determine the degree to which your items have been effective.

So now let's think in practice for a minute. So what value-- this is my question for you-- do difficulty and item discrimination indices have to your teaching? And in practice, using results from a test that you already administer, calculate these statistics-- difficulty and

item discrimination indices-- for all of it or part of the test. And just see what the results tell you.

All right. Well, that concludes module 4. And I look forward to seeing you in module 5. Thank you.

