Writing Better Multiple-Choice Questions

A Magna Online Seminar was presented on November 4, 2014 by Jim Sibley.

Writing Better Multiple-Choice Questions teaches participants to:

- Create multiple-choice questions that effectively target different cognitive levels
- Write multiple-choice questions that do not simply engage recognition processes—they actually trigger retrieval learning processes that are known to enhance later recall
- Achieve “proper construction” of multiple-choice questions
- Continually revise, revamp, and improve your multiple-choice questions

Editor’s note:

This is a written transcript of an audio recording. Our policy is to edit only the occasional unintelligible phrase. Everything else appears exactly as it was spoken.

© 2014 Magna Publications Inc.
All rights reserved. It is unlawful to duplicate, transfer, or transmit this program in any manner without written consent from Magna Publications.

The information contained in this online seminar is for professional development purposes but does not substitute for legal advice. Specific legal advice should be discussed with a professional attorney.

To make this program available to all your faculty and staff, contact Magna’s Customer Service department at 1-800-433-0499 Ext. 2 and ask about our Campus Access License.
Rob Kelly: Hello, and welcome to Magna's online seminar, Writing Better Multiple-Choice Questions, co-sponsored by Magna Publications and The Teaching Professor. I'm Rob Kelly, editor of The Teaching Professor. And I'll be the moderator today. I'm pleased you could join us. If you haven't already printed the handouts, select the file you wish to print from the file share box on the left of your screen. And then click the save to my computer button to download, open, and print it.

You may listen to this seminar through your computer, or you may choose to listen through your telephone. To listen through your phone, dial the number and use the access code shown in the box at the bottom left of your screen. And now I'm pleased to introduce our presenter, Jim Sibley. Jim Sibley is the director of the Center for Instructional Support at the Faculty of Applied Science at the University of British Columbia in Vancouver, Canada.

He is an active member of the team-based learning collaborative and is currently serving on the board. His book Getting Started With Team-Based Learning was published in May 2014. Welcome, Jim Sibley.

Jim Sibley: Thank you very much. Good morning or good afternoon, everyone. I want to talk a little bit about writing better multiple-choice questions. And I'm going to tell you a little bit about why they became very important to me. About 11 years ago, I discovered a life-changing pedagogy. And that was team-based learning because it let me do things in large classrooms that we didn't think was possible. But a real key to that was actually writing really good multiple-choice questions.

And we did the common novice mistakes of writing some fairly poor questions and rushing them to our classroom and enduring the wrath of students. And we quite quickly decided we needed to figure out a better way to do this. So what we're going to talk about today is a few things. We're going to look at the multiple-choice format overall, some of the vocabulary we use when you look at the literature on writing multiple-choice questions. We'll talk about how you write the question itself and how you write the, all the what are called options.

But we'll talk about the vocabulary in a second. We'll talk about something called Bloom's taxonomy that many people are probably familiar with. But it can be really helpful to help you write multiple-choice questions that test different levels of understanding. We'll look at some common construction errors. If you had a chance to look at the Faculty Focus article that we just published a couple of months ago, it goes through some of the common construction errors. And this is your chance, after you've written a multiple-choice question to check your work.
We're going to talk about something called item analysis. One of the nice things about writing multiple-choice questions, is you can get some feedback on how well that question is actually working. So we'll look at item analysis to identify questions that you might want to work on in the off season to make sure they're a little bit better for next year. Finally, I'm going to give you the last word, a couple of parting pieces of advice to make sure you get started on the right track.

As far as answering questions, I encourage you to keep typing questions into the chat area down at the bottom there. And as we transition through each section, I will call on David to help me work through some of the questions that you've come up with. And then there will be an opportunity at the end to ask even more questions. So when we look at multiple-choice questions, there is some vocabulary we want to think about. And we're going to just touch on that.

But first, I want to know a little bit about how you're using multiple-choice questions. And the first one is, how often do you create and use multiple-choice questions? So it's looking like a lot of us use multiple-choice questions quite often. And I think a lot of us because of larger class sizes reduced TA hours have been, we feel like we're forced into using multiple-choice questions because they're so efficient, and they can be auto-graded.

This will lead to my next question. Can you bring up the next question? What statement best describes your own belief about multiple-choice questions? Format is flawed and shouldn't be used, multiple choice questions can only test lower level understanding, or multiple choice questions can test higher level of understanding.

This one is an interesting one because my own beliefs around multiple-choice questions have evolved over the years. We can see that most of us actually think that you can get at higher-level understanding with multiple-choice questions. But I think one of the things that we often bump into is our own undergraduate experience. We've probably been exposed to really terrible multiple-choice questions. And we often will blame the format rather than a poorly written question. Rob can you bring up the next slide?

Next slide, let's see. There we are. I think we're both trying to do this. Okay. So my beliefs around multiple-choice questions have evolved. In the supplemental material, I pointed an article by Bob Bjork called Multiple-Choice Questions Exonerated At Least of Some Charges. And he points at multiple-choice questions being a really good vehicle to enhance learning. We often think that short-answer questions are the answer
because in a multiple-choice question, doesn't the student just need to recognize the right answer rather than retrieve it?

And he's done some interesting studies to actually find out that the nice thing about multiple-choice questions is you strengthen the student's understanding of all the presented options, not just the single correct retrieval. But the point he makes in that article is multiple-choice questions have a really valuable place in learning.

They might not have a valuable place in assessment all of the time. But in learning, they are really, really quite excellent. I really encourage you to look at that Bob Bjork article. I actually find all his writing really interesting. He runs The Learning and Forgetting Center at the University of California. And all of his work has sort of pithy, authentic advice at the ends of the papers. He distills it down to in the classroom, you want to think about this. And so I really love his work.

And that's in the supplemental material that will point you to there. So now I've brought up what is just a typical multiple-choice question. The question is posed. We have four choices in the case of this question. What I want to go through here is the vocabulary that when you go into the multiple-choice question writing literature that you'll bump into.

So the question at the top, the actual question part of a multiple-choice question is called the stem, or it is sometimes called the question leader. The A, B, C, D or A, B, C, D, E, that's known as the option. It will contain the correct answer and the incorrect answers. The correct answers, you'll sometimes see them referred to as keyed responses in the literature. The incorrect answers you'll see referred to as distractors or foils. For some reason, I really dislike the name foils because I don't want to trick my students and foil them, but you will find it in the literature.

So the options. And then the whole thing, the question and the stem and the options combine to be an item, which is that standalone multiple-choice question. Later on, when we talk about item analysis, we're actually talking about a whole question analysis. So if you have any questions, start typing them into the chat thing. And when we transition to the next piece, I'll get Rob to help us through if there's any questions there. When we begin to write a multiple-choice question, we have to do a couple of things.

And the first thing is we have to select important test-worthy knowledge. And I think this is a place that we sometimes fall down when we're writing multiple-choice questions. We will start writing questions without thinking about, what do the students really, really need to know? What I do with my instructors is we often encourage them to write reading guides
for the student where they'll go through a reading, and they'll write question stems. And they might write 30 or 40 or 50 of them. And then they'll look through those stems for what's really, really important.

If you jump straight to writing questions, you tend to write questions at a very low level. If you step back a little bit and identify this important test-worthy knowledge, you can start writing questions at a higher level. The next part of stem development is if you can, you want to write a standalone question. And that's an if possible. There are many examples of excellent multiple-choice questions that aren't standalone, but that's the gold standard, to try to make it a standalone question.

I should be able to put my hands over the options and successfully answer the question if it's well written. We want the stem to be grammatically complete. So that idea is you don't want to have questions that, you know, the answer is dot, dot, dot or the answer blank, and what is blank the answer. We want to get a grammatically complete question to be as clear as possible for the student.

We also want to try to avoid negative stems if possible. If we use not or negative words in the stem, you'll want to bold them or underline them. And the reason for this is you don't want to make the question tricky. You want to make sure you're testing the student's understanding of the contents, so we'll bold or underline the not statements.

I want to introduce Bloom's taxonomy because it's a really nice place to begin crafting your questions from. Benjamin Bloom was at the University of Chicago in the '50s. And he was tasked with bringing rigor to the comprehensive examination process. At that point, they didn't have any kind of taxonomy to think about levels of learning. And he developed, along with some colleagues, and published in 1956 Bloom's Taxonomy in the Cognitive Domain. And what it is, it's a layer of understanding from low level to high level starting at I can remember it. I understand it. I can apply it. I can sympathize. I can analyze. I can create with it.

So it's this progression of understanding. The lovely thing about Bloom's is there's a number of verbs that associate with every level. When you look at a multiple-choice question, there's going to be a verb in the stem. And which verb you use controls the level of the question. Now this is a short form of this table. In the supplemental material, there is a chapter excerpt from my TBL book that has a much more extensive table version of this. But the nice part about this is after I've identified test-worthy knowledge, I could select some verbs across the taxonomy to test the student's understanding at different levels.
Someone called Barton has really helped us with this. And what he developed, and again in the supplemental materials, there's a more extended version of this, he developed question leaders for each one of the Bloom's level. And they can be a really handy place to begin constructing your questions from. Once we've developed that question stem that has that Bloom's verb in it, that we know the level of the question. We hope it's standalone. We hope it doesn't have a negative statement in it. We're ready to begin to develop the choices, the options.

And there's a number of rules around the option development. One is we want to keep the options as short as possible. What you'll often find is as you write the options, a phrase will consistently appear in the options. And we usually work with that phrase and try to move it up into the stem to keep those options as short as possible. You don't want long, rambling options because you can make the question tricky just by the length of the option.

The other thing you want to consider when you're developing options is, we want to have a correct answer that's clearly and defensibly the best. And we want incorrect answers that are clearly incorrect. But the line that we're kind of straddling there is we want to make them plausible enough that we are testing the student's understanding.

In the supplemental material, I point you at the National Medical Board has a really great PDF resource called, Creating Multiple-Choice Questions for the Basic and Clinical Sciences. And it talks about this correctness to incorrectness as being a continuum. And the interesting thing about that continuum is if the correct answer is at one end, and all the other options are at the other end, it will be quite a clear question for the student.

If options start to cluster around the correct end of the spectrum, you've created a question that will likely make the student mad. It can be a good question for starting a discussion, but it might not be an appropriate question for a test. We usually try to avoid using the all or none of the above, mostly because if we go up, and we're trying higher-level understanding to test higher-level understanding, nothing is absolute once you get to higher levels, and the all or none often doesn't apply. And we can get students arguing quite vehemently for their version of how the world works.

We also want to keep the options similar length so we don't kind of give away any hints on what might be the right answer. Just like when you write lists, we want parallel construction. So if you're using all -ing verbs, you want to kind of use -ing the whole way through. We also want to talk about grammatical consistency and make sure that there isn't any cues in
that question stem that are going to help students eliminate some of the options.

And we're going to talk about that in a few minutes when we talk about construction error. The other thing is numerical options. We want to make sure we get those in numerical order, you know, 0.1, 1, 10, 100. We want to put it in an order so that it's not tricky. Because if a student scans down, and the numbers are in random order, you've just made the question tricky. You haven't made it testing the student at a higher level. So I'm just going to pause for a second, Rob. And do we have any questions that have appeared in the chat window.

Rob Kelly: Yes. We have some questions. When we ask for the best answer or choice, does that mean there are other correct answers?

Jim Sibley: So one of the things when you ask for the best answer or choice, you're going to have words in the stem like, what is the most likely outcome or what is the best thing to do. So you're giving students a criteria by which to discriminate between good options and great options. The nice thing about getting those kind of discrimination words in the stem is you can really ratchet up the difficulty and the plausibility of all the options. You want to make sure that you're using, you don't want to go, what is the answer? You want to say, what is the best answer? So you've got to have those sort of discriminating words wrapped around the question.

Rob Kelly: Okay. Nursing certification tests have A and B questions. Do you have any advice about this type of question?

Jim Sibley: I need a bit more information . . .

Rob Kelly: I think what it is there's two options.

Jim Sibley: Oh, yeah. That's usually lazy question writing. And it's also, it can be tricky, right? Because the student can motor down through the question, find a correct option. You know, put down that, you know, that value and miss the fact that further down it says A and B. So you can actually make the question tricky by doing that. And so you want to try to avoid that.

If you want to ask a question about A and B, ask a question about the relationship between A and B because then you're getting at they need to know both things. But they're just telling you, just measuring the relationship, which is, I think, a really nice workaround for those A and B questions.

Rob Kelly: Okay. And then what is the best number of options to give for a question?
Jim Sibley: This is a personal preference. The literature is pretty clear that four options is enough. But there's a lot of people that like five options because if the student is completely guessing, they don't want them to get 25%. They're happier with them getting 20%. But there is some literature that four is fine.

Rob Kelly: Okay. And there's a question about item analysis reports. And I think you'll probably get to this later on.

Jim Sibley: So, yeah, we'll get to that a little further on. We're not going to talk about the value point biserial. We will talk about the discrimination later on. For a deeper look at item analysis, there's two things. Google it. You'll get more information than you possibly need. Go talk to someone at your teaching and learning center. Find a psychometrician on campus that can work with you. Also in the supplemental material, my chapter excerpt has a bigger description of all the kinds of item analysis including point biserial and P values and all sorts of things. But we're mostly going to talk about discrimination in depth today.

Rob Kelly: Okay.

Jim Sibley: Okay. So let's go on. I want to talk about construction errors. There are some very common errors that we'll all make, you know. As we write questions, there's a human tendency to do certain things. My favorite story about construction errors is, my partner used to work for the Canadian Blood Services as a research scientist. And every six months to a year, she needed to recertify on the standard operating procedures. And so there'd be a little online module followed by a multiple-choice test that she needed to take to recertify. And I was always intrigued by these tests. So I would take the test after she had taken it just to see. I could typically get 80% to 85% on that test based on recognizing construction errors and using them to guess which is the right answer. Students, and some students are absolutely expert at this, there's actually resources for medical students to get them ready to be test-wise so they can spot these errors and benefit from them. We want to stamp these errors out if we can so that the test-wise student doesn't benefit, that we are actually testing their content knowledge, not their ability to deconstruct a flawed question.

The first one we'll often see are grammatical cues. And what happens is, you'll write a question stem. And in the stem, you'll have some, a common one is you'll have a singular stem. I'm looking for a singular answer, and you'll have a plural option. Sometimes you'll ask, you know, which verb applies? And then you'll have a noun option. The example I have here is a really simple multiple-choice question. But if you look at it, and you read
it, and you think about it, there's only one option that grammatically fits with an. It needs to be an option that starts with a vowel.

And so a test-wise student, if they don't know the answer will use that grammar cue to guess more effectively. Another one we often see, and when I talked about option development, we talked about we should try to keep them the same length. And the reason we want to keep them the same length, it's a rule that's sometimes called too long to be wrong. And what happens is you'll write a series of options, the correct answer and some incorrect answers, and you'll feel compelled to write a little more detail around the correct option so it really truly is correct.

And a test-wise student will pick up on this. And if one of the options is longer than the others, they'll guess that one, hoping that you have fallen into this trap. Another place we can get ourselves into trouble is we can have logic cues, and it's called overlapping options. If you think of a question that has ranges, and if those ranges happen to overlap, and the students have been working with a variety of background materials, they may have different ranges in answers that they think are correct. And they'll argue vehemently when you get this wrong, and the example will show you better.

I'll give you a moment to read the question. Now when you look at those options, you know, if some of their reading material had, you know, it's around four hours, you know, suddenly A and B could be correct. Or, you know, the range is 6 to 11 hours, and suddenly it gets really messy about which one is exactly the best answer. In this case, you'd probably want to try to not make the option or the overlapping ranges so close. And you would put some words in the stem that might say what's the most common range or what's, you know, you want to get some discriminating words in the stem to help students pick a range that's reasonable.

Repeating words is another one we often do. We'll take a phrase or a word from the stem, and we'll reuse it in one of the options. This is especially true in questions that have really difficult vocabulary. If I see the same phrase or the same start of the word in one of the options in the stem, I will likely guess that option. And here's the example for that. Really simple looking question, but the fact that the question stem and option D contain the same word are a bit of a cue to me that that's probably the right answer.

This would be a much better question if in option D, we just got rid of nuclear. We talked earlier about avoiding absolute terms. A lot of this has to do with, you know, higher-level ideas are not absolute. You can use some absolute terms down at lower levels of understanding. But as we move higher up the Bloom's taxonomy, absolutes become much more
dangerous to use. And the solution with absolute terms is to make sure those words in the stem like, the most likely, the best, the least likely that you give students a criteria by which to sort through plausible options to select the best answer.

There's another human tendency we have as we develop our multiple-choice questions, and you've written, you know, maybe 20 for your test. If you actually go back and reflect on where you've put the correct answer most of the time, it's often C or D. It's often C in a four-distractor question and D in a five-distractor question. So it is worth checking your work after you've written the question to go back and look where you have put the correct answer. This is especially true when we write numerical questions with a numerical list.

For some reason, we often don't like to make the answer the first one or the last one because that's too obvious. One last thing about questions, and you can Google this, Texas two-step multiple-choice questions. It's a way to turn a multiple-choice question into a much higher-level question really simply. And how you do it is you take the verb that is in the question stem.

It might go like this. You have a question that asks you to describe X. You change that into what is the best description. And then you have a number of plausible descriptions. And the students need to discriminate. You have pushed that question from describe, which is a fairly low level understanding up to a discrimination question which is much higher. This is a really interesting way to increase the level of your question. So, Rob, how are we doing with question . . . area.

Rob Kelly: We have a couple questions here. Can we rely on answer randomization by software to avoid too many C and D answers?

Jim Sibley: Yeah, there's a number of solutions for that. There is a piece of software called Test Maker from Epstein Education. I'll touch on that right at the end. And it will help you randomize where the right answers are. I think you can probably, this is a case of you just need to check your work. But I'll point you at a way that you can randomize them with this software at the very end of today. Any other questions?

Rob Kelly: Doesn't using a lot of adjectives best, most likely, least likely make the answer a bit subjective?

Jim Sibley: Well, there's kind of two pieces there. Students who are living in the black and white world, there's only one correct answer, don't like the adjective stuff. As we move students up into discrimination, you know, because what would be the best management plan for this complex situation is a
great question. But if a student is at a low level of something called Perry's Framework, they're going to see the world as black and white. And they're going to want a really hard answer. So I think straying into the best, most likely, least likely might depend on the level of your learners.

If you're teaching a first year course where people still think that the disciplines are black and white, and there is a correct answer, that can be a bit more dangerous. But as you go further in the discipline, and you start to realize every discipline is messy and gray at the top end, those best, most likely, least likely start to make a lot of sense.

Rob Kelly: Okay. One more question. Have you employed tables in Microsoft Word to rearrange options?

Jim Sibley: I have not. We can talk offline and figure out what that is about because I use these scratch cards for multiple-choice tests. And I'll show you those at the end. And basically, the options are already printed on them. And I actually need to rearrange my questions to match the card. And that's the way I typically get randomization because I'm matching a randomized key. I'll show you those at the end. But if there's a way to do it in a table to sort of check your work, that would be great.

The big message here is you've got to check your work. Writing multiple choice questions, good questions is about checking your work a lot. And it's also about having other people check your work if you can.

Rob Kelly: Okay. There was one more question. Would you recommend different values for different question difficulty?

Jim Sibley: Yeah. You could make an argument that that's a reasonable thing. Because if I write a lower-level question that a student can answer in a minute, and I write a higher-level question that takes the student three or four minutes to work their way through, it actually might make some sense to value those differently.

Rob Kelly: Okay. We still have some questions coming in. I don't know if you want to take those now, or would you like to move on?

Jim Sibley: Sure.

Rob Kelly: Okay. Do you ever have someone else check your work? Sometimes I'm too familiar with the questions, and I don't see my errors.

Jim Sibley: Well, I mean, that's a great point. I always have someone else check my work. Because, you know, I'll gloss over the missing word or the ambiguous statement that I've started to understand in my own head, and
it's amazing the number of times I give the question to someone else, and they pick the wrong answer. So I think it's really important to find someone else that can check your work. And there's a couple ways to do this. If you have a TA, get it in front of a TA. Have them work with it. If you have a colleague, work with them. They can write multiple-choice questions. You can write multiple-choice questions. I think the most interesting one I heard about this was Holly Bender is at the University of Iowa State. She has a lunchtime session where people bring their multiple-choice question, write them up on flip charts, and they try to pick the best multiple-choice question each day. And what it requires people to do is go around and really dig into what the question is and how it works. And they do this over lunch once a week. And it's a way to kind of keep working on your questions.

Rob Kelly: Okay. Do you want to take some more questions?

Jim Sibley: Sure.

Rob Kelly: Okay. For higher-level questions, how do you determine how long to give students to answer?

Jim Sibley: We typically give one minute per multiple-choice question on a lot of our tests that we do. Our tests are thoughtfully constructed, however, at remember, understand, and light application for where we're using the test in the learning cycle. If you were asking bigger, more difficult questions, you might want to give more than that one minute.

Rob Kelly: Okay. And what are your feelings about the select all that apply questions?

Jim Sibley: It's usually a lazy question writing strategy. But a lot of people like them because they can go through a lot of content quickly with the questions. But you don't often see select all that apply that are at a higher level. They tend to kind of be down at the understand level. So if you end up writing a select all that apply question, I would challenge you to look at which Bloom's verb is involved in the stem and consider whether it's doing what you want it to do because sometimes we do want to test understanding, but we don't want to test just understanding.

Rob Kelly: What is your opinion about K-type questions?

Jim Sibley: I'm going to sidestep this question and point you to the supplemental resource. The National Medical Board PDF that I talked about has a long discussion about K-type questions. A lot of people really like them. Because of where we use multiple-choice questions in the learning cycle, we don't use K questions much. Okay. I'm going to go on. But keep typing
questions into the chat area, and we'll have another opportunity to look at those in a few minutes.

Now I want to talk about item analysis. So you've built a really great test. You've taken it to your classroom. You've given it to your students. The nice thing about multiple-choice questions is you're going to get some statistics out of the backend on how the students answered those questions. And we can use this item analysis to help us identify questions that aren't working the way we want. What we're going to look at mostly here is something called discrimination.

Again, I'll point you to the supplemental materials the chapter excerpt has more information on all the item analysis variables, not just to discrimination. But I want to get you thinking about this discrimination value. Basically, discrimination helps us determine whether the students that prepared well and got a good score on a test did better on a question than the students that didn't prepare as well and got a lower score on the test.

And what we're usually hoping is for a positive discrimination, which would lead us to believe that students who did better did better on the question than students who didn't study, did poor on the question. Contrast this to a negative discrimination, students who didn't study for some reason are doing better on this question. This will make more sense in a second when we look at the discrimination table.

But these negative discriminations are red flags that we've done something wrong. And we probably want to fix that question. Now this is a big scary table. But I'll walk you through some of the values here. And I'll back up a little bit. One of the things with discrimination and item analysis and the whole thing is it's not necessary that you know how to do this. It's necessary that you know someone on campus that can help you with this. I have faculty members who bring all their test results to me every summer. We look through. And I say, you know, if you only fix two questions on this test, fix these ones. I would encourage you to find a psychometrician or someone at your teaching and learning center to help you with this. But I'll walk you through some of this. If we look at the row for question one, in the center column is labeled discrimination, you can see that it's a positive value of 0.46. And that's a good value for discrimination.

If we look to the left of that at the percent correct which is three columns, whole group, upper 25, lower 25, what discrimination is actually doing is comparing the upper quartile of students that the people that did really well on the entire test against the people that did poorly on the entire test and their performance on that particular question.
What we're typically wanting to see is that the people that did really well on the test do well on this question. And if you look at those numbers, the whole group, 81% of the students got the question right. In that upper quartile, 97% of the students got it right. And in the lower quartile, 56% got it right. So that question seems to be discriminating between students who prepared and students that didn't. Contrast that to if we go down to row seven, you can see that negative discrimination. And that should be a red flag that there's probably something wrong with this question.

And if we look at the percent correct columns there, you can see that 55% of the whole class got that question right. Fifty-three percent of the upper quartile got it right, and somehow 59% of the lower quartile got it right. So the less prepared students did better on that question, which probably means there's something wrong with the question. You'll see very negative discriminations when you've made a mistake, and you've keyed the question wrong. You know, the correct answer is B, but for some reason, you wrote down C, you'll see these negative discriminations pop out.

The nice thing about this discrimination column is that I can quickly scan down that, and I would, you know, question seven, we absolutely with a negative discrimination will have to look at that in the summer. And if you look at question three, it has that lower discrimination of 0.17. We're going to want to look at that. The thing about discrimination is it gives you inferences to ask the question, is the question doing what you want it to do? There are some questions where, you know, you want 100% of students to pick the right answer, and you will get a discrimination value of zero.

The question might be doing what you want it to do. But seeing that zero would send you back to the question to check. I've touched on this. The big thing with item analysis though is not to get hung up with understanding all the statistics. Try to get someone else to do that for you. And really target your time because I'll have faculty members come to me and say, I have time to fix ten questions. Point me at the ten questions that I should fix. So can we, Rob, can we go back, or, David, can we go back? And are there any more questions.

Rob Kelly: Yeah. There's, I'm not sure if I understand this one, but I'll ask it. How can low achieving students score higher on a hot question as compared to higher achieving students?

Jim Sibley: I'm not sure what a hot question is. But in that case where you actually just say selected the wrong correct answer, you know, you keyed it wrong, you will often set it up so that, you know, the low-achieving students are picking the common misconception, and they're getting the answer right
because you keyed it wrong. That's very often what you see with the negative discrimination. The other one is if it's slightly negative, it often points that we've written a trickier question. For some reason, we've made it tricky and difficult to understand. And again, it will just point me back at the question to just think about whether I should improve it.

Rob Kelly: Okay. And that was, that HOT is an acronym for higher-order thinking. I didn't pick up on that.

Jim Sibley: Okay. I think my answer still applies.

Rob Kelly: Okay. Do you have a rule of thumb for what percentage of students should get an item correct?

Jim Sibley: Sure. Again, in the supplemental material, it talks about something called P-value, which is really the percentage of students that are getting something correct. And you're typically looking for a range between 25% and 75% on a particular question. You want a range, you know, 25% being a hard question, 75% being an easy question. But you're typically inside that range. And you can look at this thing called P-value to make sure that your test is performing the way that you want it to.

Rob Kelly: Okay. If you get a very negative discrimination value, should you still include that question as counting for that testing group?

Jim Sibley: There's a variety of ways and places to go from here. When we get a very negative discrimination, i.e., we've typically screwed up, we will often take that question out of the test. And in some courses, you'll even see, you know, the 20-question test of last year becomes an 18-question test next year because they've removed the questions that are problematic.

Rob Kelly: Okay. Can a class with mixed abilities reflect on item analysis results not how a stem or question is phrased?

Jim Sibley: Ouch. I'm trying to, a class with mixed abilities. I'm not sure I understand the question.

Rob Kelly: I think what it's asking is students with varying abilities in a class, would that show up in the item analysis and not necessarily reflect a problem with the question?

Jim Sibley: Well, okay, great question then. What this will point out is we should be very careful in asking really clear questions because if we think of a mixed ability class, some students may have lower levels of literacy or more difficulty in test-taking situations. And we've got to help them by asking
simple, clear questions that get a content understanding, not that they could wade through a big complicated question.

Rob Kelly: Okay. A couple people asked about true/false questions. And you did talk about A, B questions. And I don't know if those comments would apply to the true/false as well.

Jim Sibley: So a lot of people do like the true/false questions. They can be relatively useful for getting through some lower levels of understanding. They're not very useful for higher levels of understanding. But you'll see people use them down at the lower Bloom's levels, remember and understand. I tend to avoid them because I usually want to try to push my students up a little higher in Bloom's level.

Rob Kelly: Okay. Is there a threshold value for positive discrimination if you are intending to discriminate between the less and more prepared students?

Jim Sibley: So if you go out on that Internet that's ever, ever wrong, and you ask this question, you'll see that a threshold value of around 0.4 is often bandied about as if the discrimination is above that, it's a good question. If the discrimination is below that, it's a bad question. That has not been my experience. I use discrimination to point me at looking at the questions and thinking about whether they're doing what they should.

Rob Kelly: Okay. Should we curve, for example, if the highest grade is 90%?

Jim Sibley: Well, two things here. You may want to curve in the first iteration of a test, but you'll want to look at how difficult that test is for subsequent iterations. Okay. I'm going to, I'm cognizant of the time. I just want to move on and leave you with a few things to think about as you start writing better multiple-choice questions. Start with Bloom's. You'll have Bloom's taxonomy and the Barton question leaders and the Bloom's verbs tables, great place to start.

Spend the time. This is the greatest gift you can give yourself. We often reel off a set of multiple-choice questions in short order, inflict them on our students, and there's nothing more uncomfortable than standing in the class and taking all that in. Have a talk with yourself, and make sure you spend the time to write good questions. Now there's good news and bad news here. You can write good questions with effort. The bad news is the better you get at writing multiple-choice questions, typically the longer it takes you to write a question.

And I think the reason for that is your standards go up so much. The other piece is never write questions alone. Making sure that you've got a colleague or a TA or someone to read over your questions to get rid of
some of those common ambiguities that you don't see anymore. Check your questions for flaws. So we've got to check our work. And even better, get someone else to check your work. And then use that item analysis to improve your question. Find someone on campus that can help you get through item analysis and point you at improving your questions.

I do want to touch on there's these really wonderful, they're scratch and win multiple-choice cards. They're called Immediate Feedback Assessment Technique. They're available from Epstein Education. We give our multiple-choice tests individually. And then we have students get together in teams. And they use these scratch cards. And basically it's like a lottery scratcher, and you scratch off looking for the right answer. But there's incredible conversations that these foster as, you know, we will go around the group and say, what did you get, what did you get, what did you get?

If it's D, and we scratch off, and we get it right, right away, we're never talking about that. But it gets more interesting when there's disagreement. And we scratch off, and we get that immediate feedback to go back into the question and talk about it some more. And we push students back into the question with these scratch cards by using decremental scoring so they get four points for the first scratch, two points for the second scratch and one for the third. These are so much fun.

And I see that the slide is not in the handout. That is true. I'll get everyone's e-mail. And I'll send a link to a page that talks about these on my website. So final few questions. Yeah, Susan pointed out that in my description of item analysis, I was making the assumption that a less prepared student generates a lower mark. And that absolutely is an assumption. But it is in the world of kind of the way we look at it. Item analysis isn't about you are more prepared or you are less prepared. It's about me helping look into my questions for where I might want to spend some effort improving my questions.

I also see a question there about do you give exams back? People often protect their multiple-choice questions quite dearly. Because they're so long to create good ones, we often don't give back the multiple-choice question portion of an exam. And we'll often do things like if the students are working in team, and there's five team members, if there's not five test sheets in the folder at the end of the class, everybody gets zero. We really hang onto our questions tight.

And I just want to point you at my website which is learntbl.ca. You can go there, and there's resources on multiple-choice question writing. And it really talks about my first love, team-based learning, in more detail. And there's a page there that talks about those scratch cards that you can figure
out where to get those from. I see we're at the top of the hour. David, do you have any final words?

Rob Kelly: Yes. And, actually, there are some leftover questions, Jim. And I will forward those to you. And we will send your responses to all who attended today.

Jim Sibley: And I'd encourage anyone that has a question, type it into the chat right now or the minutes afterwards. And then I will answer all those in an e-mail to everyone.

Rob Kelly: Great. Thank you, Jim Sibley. And thank you all for joining us today. As a reminder, in about an hour, your campus will receive an e-mail with the link for this presentation's recording, which will be available for viewing the next 30 days. And we will be mailing your CD near the end of that timeframe. Your campus has received an e-mail evaluation form from us. Please fill it out and tell us what you think of today's program and what programs you'd like to see in the future. Complete information about our upcoming seminars is available at www.magnapubs.com. Thanks again for joining us, and have a great day.

Follow-up Questions & Answers

Reading Guides

Florida International University: Do you give students the reading guides? If so, doesn't this defeat value of student attending class?

We use reading guides to focus student attention during the required out of class reading. Since our classroom sessions are focused on using what we learned in the readings to solve interesting problems in class, the reading guides are a must. We don't cover the readings in class. We start where the readings leave off. Student must be seeing value in coming to class prepared. Our attendance rates in large second year engineering courses hovers around 95%. This is the power of doing great activities and the power of accountability to ones teammate peers. The fact that your teammates notice when you don't come to class and feel they do more poorly without “all hands on deck” unleashes some subtle and not so subtle peer pressure to come to class prepared and contribute to your teams efforts.

Opinion about K type questions
Louis Kutcher: Nursing certification tests have "a and b" questions. Do you have advice about this type of question?

Michael: What is your opinion about K type questions?

"K questions" refer to the...A and B, A and C or B and C type questions. Some people really like these kind of questions, but I find them too tricky. I can feel myself tense up when I get one of these and try to really concentrate to make sure I understand the question and those complicated choices. I guess the question is....are you seeing how well the student read the question or how well they know the content?

**Using Subjective words/adjectives in question stem**

Ms. Shazia Khalid: Doesn't using a lot of adjectives (best, most likely, least likely) make the answer a bit subjective?

*Whether adjectives are appropriate, depends on the Bloom’s level of the question. With higher-level questions we are often discriminating between reasonable courses of action. With a low level “can you remember this” question, the answer is more black and white – you shouldn’t use adjectives in these cases.*

Ms. Shazia Khalid: when we ask for the best answer/choice, does that mean there are other correct choices too?

*We might want to think of making other choices plausible, but still incorrect. The concept of best answer bring to mind the value of placing the options on a continuum from correct to incorrect, with the correct or best answer....clearly the best answer and the incorrect options clustered away from correct end of spectrum. The other nice part of thinking about putting options on a continuum is that it reminds us to build options that are related. We don’t want to include that one throw away option from outer space, just because we couldn’t come up with a good final option.*

**Other stats in Item Analysis**

Michael: I have seen some item analysis reports that use point biserial and others the discrimination index. Can you discuss the merits or pitfalls for each?

*I would refer you the supplemental material for explanations of other item analysis stats. Find someone at your local teaching and learning centre to help with this. Google it....you will get tons to read.*

**Opinion on All of the Above Questions**

Southwestern Illinois College (SWIC): Can't "all of the above" options be useful

*These can be satisfying questions to write because you get the cover many pieces of content in one question, but you often are not really asking higher level, important ideas. If you are all about covering all the content...theses questions can be satisfying.*
Barton Bloom’s Question Leaders Reference
William Carr: Barton Reference?

There is a reference to Barton's article in the Jim Sibley's book chapter excerpted in the supplemental material. (David Rice, Magna Publications)

Quality (or lack there of) in Test Bank Questions
St. Margaret's Episcopal School: SUNY, in my experience, absolutely not. Most/many software banks are terrible

I completely agree test bank question are often terrible, but they can be fodder for writing better questions from.

Time allowed per question
William Carr: For higher-level questions how do you determine how long to give students for that question?

St. Margaret's Episcopal School: are questions that respond to text, sources, etc., of greater value/difficulty?

We usually give one minute per question for individual tests and 1.5 minutes per question for team tests. My MCQ tests are specifically designed at remember/understand/light application...if you ask higher-level questions you will likely need to give more time.

Rule of thumb on percentage of students getting question correct
SUNY College at Brockport: Do you have a rule of thumb for what percentage of students should get an item correct?

Percent correct corresponds to something known as P value...we typically aim for P values between .25 and .75. A .25 P value question means 25% of students got 1 question write – a harder question. A .75 P value question means 75% of students got it right – an easier question. We want a mix.

Response to very negative discrimination
University of Waterloo 4: If you get a very negative discrimination value, should you still include that question as counting for that testing group?

A very negative discrimination in my experience is often my fault. Typical, I have missed keyed the question. Whether you keep the question may depend on context, but I am normally very willing to throw out poor questions to be fair to students.

Opinion about True/False questions
Susan: Do you think there is a place for True/False questions?
They are only really useful for very low level questions – do you remember this? They are often not even useful for the next level up – do you understand this? Depend on the goals of the test they can be appropriate, but I never use them myself.

Is there a universal positive discrimination threshold value?
University of Waterloo 10: Is there a threshold value for positive discrimination - if you are intending to discriminate between the less/more prepared students?

There is no specific threshold value that can be universally applied. The discrimination just gives you some information on how well the question is discriminating. Some easy question...everyone might get right....so low discrimination. In harder questions you are hoping for discrimination values that indicate you actually discriminating between high test score students and low test score students. We make the assumption that a high test score equates to a prepared student and a low test score equates to a less prepared student. This may or may not be the case.

Keeping your question secure
University of waterloo 6: do you give your exams back?

We are very careful to collect the MCQ tests, so they don’t end up over at the frat houses. You spend so much time building good questions it is worth guarding them. We hand out tests in team folders and let teams know that if a test is missing from the folder when they hand it in...the whole team gets zero. We have some tests being successfully used in their 10th year.

IF-AT aka Scratch Cards
St. Margaret’s Episcopal School: this slide is not in the handout...can it be provided!?

Southwestern Illinois College (SWIC): Do you use the scratch off cards as test prep, or for use during the test?

Check out the IF-AT page on my website
http://learntbl.ca/if-at-immediate-feedback-assessment-technique/

ESL issues
Cindy Pierce: Is there any specific literature addressing writing questions for students who are ESL?

I don’t know on any specific literature...I am sure there is some – I just not familiar. But this is a call to write clear questions with clear options. Write high quality questions and you will serve the ESL student well.
Adobe Connect chat:

SUNY College at Brockport: Very often, weekly for some classes.
Lori Hart: C does require significantly more work!
Red Deer College: You should separate this question into create (seldom) and use (often)
St. Margaret's Episcopal School: Higher level content understanding...application is another matter entirely, imo
  Ms. Shazia Khalid: they can present a challenge to students
Florida International University: Do you give students the reading guides? If so, doesn't this defeat value of student attending class?
  Ms. Shazia Khalid: when we ask for the best answer/choice, does that mean there are other correct choices too?
Louis Kutcher: Nursing certification tests have "a and b" questions. Do you have advice about this type of question?
  Michael: I have seen some item analysis reports that use point biserial and others the discrimination index. Can you discuss the merits or pitfalls for each?
Southwestern Illinois College (SWIC): Can't "all of the above" options be useful (more useful than a combo such as "A and B only")?
  SUNY College at Brockport: What is the best number of options to give 4 or 5?
William Carr: Where is a copy of Barton's article in the supplementary material?
University of Waterloo 10: Barton's isn't in the supplemental but the reference is.
David Rice, Magna Publications: William Carr: There is a reference to Barton's article in the Jim Sibley's book chapter excerpted in the supplemental material.
  SUNY College at Brockport: Can we rely on answer randomization by software to avoid too many C, D answers?
  Ms. Shazia Khalid: Doesn't using a lot of adjectives (best, most likely, least likely) make the answer a bit subjective?
St. Margaret's Episcopal School: SUNY, in my experience, absolutely not. Most/many software banks are terrible
  St. Margaret's Episcopal School: Have you employed tables in microsoft word to rearrange options?
  SUNY College at Brockport: Would you recommend different values for different question difficulty?
  St. Margaret's Episcopal School: It's fantastic! So easy to rearrange options...
  St. Margaret's Episcopal School: ...avoid C/D repetition
  Southwestern Illinois College (SWIC): Do you ever have someone else check your work?
  Sometimes I'm too familiar with the questions and I don't see my errors
  St. Margaret's Episcopal School: are questions that respond to text, sources, etc., of greater value/difficulty?
  William Carr: For higher level questions how do you determine how long to give students for that question?
  Snow College: What are your feelings about the "select all that apply" questions?
  Michael: What is your opinion about K type questions?
  Ms. Shazia Khalid: How can low achieving students score higher on a HOT question as compared to higher achieving students?
  Ms. Shazia Khalid: Higher order thinking question
SUNY College at Brockport: Do you have a rule of thumb for what percentage of students should get an item correct?
University of Waterloo 4: If you get a very negative discrimination value, should you still include that question as counting for that testing group?
Ms. Shazia Khalid: Thank you
Maria Zaheer: Can a class with mixed abilities reflect on item analysis result not how a stem or question is phrased?
Tasneem 2: What about True/False Questions?
Susan: Do you think there is a place for True/False questions?
University of Waterloo 10: Is there a threshold value for positive discrimination - if you are intending to discriminate between the less/more prepared students?
Maria Zaheer: Yes that’s correct, Thank you!
Florida International University: Should we curve, for example, if the highest grade is a 90%?
Susan: Assumption that less prepared students = lower marks and that more prepared students = higher marks. Maybe this is not always so?
University of Waterloo 6: do you give your exams back?
St. Margaret's Episcopal School: this slide is not in the handout...can it be provided!?
Cindy Pierce: Is there any specific literature addressing writing questions for students who are ESL?
South Dakota State University: What is your opinion on partial credit for questions where there are multiple correct answers
SUNY College at Brockport: Thanks so much from us here at Brockport.
Ms. Shazia Khalid: Thank you for an informative session.
University of Waterloo 6: thank you
Tasneem 2: Thank you
Cindy Pierce: Thank you.
Shelly Gooden: Thank you!
Maria Zaheer: Interesting Seminar, thank you
Cristina Lammers: thank you so much
St. Margaret's Episcopal School: thanks!
National University: Thank you
WNE COP: Nicely done
Helen Chen: well done. Thanks!
Susan: Helpful - thanks
Nathan: thanks
guest: Thank you!
Geno Spatafore: Thanks!!
Southwestern Illinois College (SWIC): Do you use the scratch off cards as test prep, or for use during the test?