Writing Quality Multiple Choice Items to Test More than Recall

Date: October 20, 2011
Location: MSU Union Gold Rooms A& B
Speaker: Kathy Doig, Associate Dean, College of Natural Science, MSU
Sponsoring Program: Lilly Teaching Seminar Series
Core Competency Area: Assessment of Learning – Testing

Introduction
Kathy Doig lead “Writing Quality Multiple Choice Items to Test More than Recall,” a Lilly Teaching Seminar, to discuss the development of examination techniques that incorporate both content and cognitive levels and also evaluate the test questions to ensure fair grading. The seminar was structured to provide background and the context for testing, best practices in test construction, and application of item analysis for scoring and test improvement. The participants were able to critique test questions and evaluate a real item analysis.

Reflections
I was particularly excited to attend this seminar to learn how to use multiple choice tests to teach critical thinking. As a science student myself, I have taken a fair share of multiple choice tests in my time and one trouble with my own science education at undergraduate level is that I felt that the testing schemes and grading system didn’t teach students to think like scientists. I learned science as a matter of right and wrong. This approach leaves little room for interpretation and doesn’t teach students how to apply the scientific method in order to address problems objectively. It doesn’t teach them to be curious about the world in which they live and seek the explanations about how things work themselves.

One of my teaching objectives is to educate students in the application and power of the scientific method rather than just scientific “facts.” But, how can a professor reach students, teach students to think this way, train them to be scientists, when he or she has a class of over 100? This is where I always came up short. Ok, test critical thinking through open-ended short answer and essay questions? Realistically, it’s impractical to expect an instructor to personally score large quantities of tests (to ensure fair grading). So, multiple choice tests are a practical examination technique to apply and are the go-to for introductory science education.

Using multiple choice questions to test recall and using multiple choice questions to test higher order thinking, however, are very different tasks. I wish my undergraduate science instructors had taken this seminar! While students, particularly introductory students, need the context to before they can use their own judgment to formulate conclusions using critical thinking, Kathy, an experienced test writer herself, made it very clear that testing for both content and cognitive levels can be done. To educate the next generation of scientists, I plan to apply the methods and suggestions discussed below:

Grading
Grading can be with an absolute referenced system (i.e., standard scale) or a norm referenced system (i.e., norm-based). As a student, I think I always assumed that I was better off with grading on a curve. But in reality, there are some strong disadvantages for using norm referenced grading. Firstly, it assumes that student ability is normally distributed, which may not be the case. As a result, students with similar scores can get drastically different grades. It also
promotes competition, rather than collaboration, between students. With an absolute referenced system, on the other hand, any student has the opportunity to get an A. This promotes study groups, discussions, and good-will between students. It also is an automatic evaluation of teaching ability or testing writing. If a test is written to material that the students should be able to process, and all the students test poorly, this is an immediate reflection on the instructor’s ability to convey the information appropriately.

I can envision situations where grading on a curve would be useful, but based on Kathy’s discussion, I am more inclined to use absolute grading as a standard practice because it promotes collaboration between students and is also an immediate assessment of the quality of instruction. Regardless of the method, Kathy emphasized that grading practices should be transparent, fair, valid, and reliable.

**Testing**

According to Kathy, any true testing situation should be an evaluation of achievement. A test question is a stimulus to elicit the correct response from students who know the answer and an incorrect response from those who do not. A test isn’t reliable if a student can get the right answer for the wrong reason so it’s important to not provide clues (e.g., spelling or grammar errors provide a hint) to the right answer.

Depending on their construction, multiple choice questions can test knowledge and comprehension (i.e., recall), application and analysis, and synthesis and evaluation. It is important to remember to test with congruence to instruction. If stated objectives and instruction are for concepts only, students cannot be expected to test on a higher level of cognition.

To maximize the ability of a multiple choice question to evaluate achievement, Kathy suggested 1) testing only important concepts (i.e., avoid trivia) and one concept per question; 2) minimizing the likelihood that students will guess correctly by using concise stems and options with good spelling and grammar; 3) avoiding irrelevant information to avoid confusion and increase speed of reading; 4) using the single best answer format to pinpoint holes in student knowledge; 5) using questions of moderate difficulty to maximize discrimination. Kathy recommends teaching to stated objectives and writing test questions based on those specific objectives (she even labels each of her test questions by the objective in her answer keys!).

The basic root of a multiple choice question is the stem (i.e., the question). To write a good stem, Kathy suggested: 1) keeping it simple and clear; 2) put as much information in the stem as possible to keep the options short; 3) avoid absolutes, vague terms, negatively stated stems, and “window dressing” (i.e., teaching in the question). “The cover-up test” is a method to assess a stem; a student should be able to generate an answer to the question before even looking at the options. To test higher levels of cognition, Kathy recommended using story problem stems which ask students to conduct some synthetic/analytical thinking.

**Item Analysis**

Item analysis is statistics on questions individually and on the multiple choice exam as a whole to inform the instructor of problem questions (ones that may need revision) and the distribution of student responses. It can help an instructor decide if multiple answers for a question should be accepted or if a question should be thrown out. Item statistics include: index of difficulty (i.e., was the question easy or difficult for students to answer; expressed as a percentage of students answering correctly; easy questions have a high index ≈90, hard have a low index <40; best questions are between 40 and 60) and index of discrimination (i.e., how well
does the question discriminate between students who know the information and those that don’t; the difference between the percentage of high achieving students who got an item correct and the percentage of low achieving students who got the item correct; >20 is good).

Kathy recommended considering score adjustments when questions had a difficulty index of less than 50 or a discrimination index of below 10. Then to improve these questions for future use, she suggested examining why the incorrect answer for difficult questions was attractive to high achieving students: Was the question confusing? Was it not an important concept?

Conclusions

I found this seminar to be a very valuable resource, particularly in improving my appreciation for multiple choice testing methods. Though it is surely more difficult to write multiple choice questions that test higher cognition levels effectively, I am happy to learn that it can be done. In teaching science courses, I believe it is important to test not only for content but also for critical thinking. I aspire to teach students how to conduct scientific research and apply it to the broader context of societal need. To do this, they need to learn to analyze and synthesize information.

I was concerned that the methods to teach these skills could only really be conducted in small, upper-level classes. Thankfully, this seminar has taught me that they can be applied in any context of multiple choice testing. As a scientist, this is surely reassuring to know that these higher level cognition abilities can be taught and tested in any classroom setting. I plan to apply these methods and I hope others will too!