ONLINE ARTICLE

Using a Reciprocal Teaching Strategy

TO CREATE MULTIPLE-CHOICE EXAM QUESTIONS

Murray Jensen

Irene Duranczyk

Susan Staats

Randy Moore

Jay Hatch

Chas Somdahl

Classroom testing is a chore for both teachers and students, yet it remains at the core of most educational experiences. In a perfect world, students would not need to be tested because they would be so engrossed in the topic that they would learn, for example, osmosis and diffusion as avidly as many students learn the intricacies of a new video game. In some educational institutions, testing and grading are eliminated (see, for example, Neil's *Summerhill*, 1960) and those schools sometimes produce positive results. However, schools that achieve high standards without exams are frequently associated with strict selection processes that admit only the best and brightest students. Additionally, schools that do not use exams are typically not supported by money from taxpayers and legislators, who demand more and more accountability (e.g., increased testing).

Teachers give tests for many reasons, but three stand out: 1) to promote and motivate student learning, 2) to produce data for evaluations (e.g., grades), and 3) to provide evidence to our administrators, financial supporters, and parents that students are learning.

This paper describes and evaluates a new type of multiple-choice test question that is relatively easy to construct and challenges students' understandings of biological concepts.

Murray Jensen (msjensen@umn.edu), Randy Moore, and Jay Hatch are H.T. Morse Alumni Distinguished Teaching Professors. Irene Duranczyk and Susan Staats are Assistant Professors, and Chas Somdahl is a Curriculum Specialist; all in General College at the University of Minnesota, Minneapolis, MN.

Common Types of Test Questions

An unwritten rule in teaching is that instructors can either spend time constructing an exam or correcting an exam—but hopefully not both. For example, essay exams are easy to create, but frequently difficult to evaluate. "Explain the membrane events involved in the formation and propagation of an action potential" would be a very easy question to put on an exam, and would indeed test students' understandings of neurophysiology. However, the effort required to correct 100, 200, or maybe even 500 responses to this question would be quite daunting, if not logistically impractical and unfair. Essay exams can be excellent testing tools that are frequently used in smaller classes where the instructor can carefully evaluate responses, however, these kinds of exams are often not feasible in large classes.

Multiple-choice questions frequently get a bad reputation from those who have the luxury and time to administer essay exams. "Multiple-guess," "promotes poor study-skills," "promotes test-taking-skills over learning," and "a sign of a lazy teacher" are just a few of the comments that are sometimes attributed to multiple-choice tests and the instructors who use them. Adding to the controversy of multiple-choice exams is their widespread use, and occasional misuse, within large standardized tests like the ACT and SAT (Moore et. al., 2002). Despite a historically poor reputation, multiple-choice exams are commonly used in large enrollment classes because they are easy to administer and correct. Additionally, multiple-choice exams can be just as valid and reliable a testing instrument as an essay exam if the questions and distracters are constructed carefully. (For an overview of testing and curriculum design, see Smith & Ragan, 1999.) The key is that the instructor

must spend time "up-front" carefully creating challenging questions.

Bloom's Taxonomy: A Tool To Help Construct Test Questions

In 1956 Benjamin Bloom developed a taxonomy of cognitive tasks (Bloom, 1956). His taxonomy involved several categories (knowledge, comprehension, applications, analysis, synthesis, and evaluation) that form a framework still used by many teachers to guide academic endeavors (e.g., the creation of learning objectives, the construction of test questions). The categories are ranked from least complex (knowledge) to most complex (evaluation) in terms of cognitive requirements. For example, knowledge tasks, or knowledge questions, include "Where in a cell are the chromosomes located?" or "What are the four nitrogenous bases in DNA?" Evaluation is at the highest level of the hierarchy and involves tasks such as "Read Watson and Crick's original paper about the structure of DNA and critique its conclusions in light of the current research on nucleic acids." (Note: Some educators use the top three levels [analysis, synthesis, and evaluation] as the requirements for "higher-order thinking skills.") Good teachers design their curriculum so that there is a balance of objectives between the different levels of Bloom's taxonomy. Ideally, there is a similar balance of questions on exams.

Entry-level biology courses have traditionally emphasized the knowledge level of Bloom's taxonomy. This is understandable, because the quantity of knowledge ("facts") in entry-level biology is enormous. Biology texts read more and more like dictionaries, and a "good biology student" is at times identified as a student who can memorize the most biological facts. (Note: Most scientific policy documents strongly advocate "depth over breadth," meaning that instructors should be teaching fewer concepts, but strive for more in-depth understanding with each. See, for example, American Association for the Advancement of Science, 1993, and National Research Council, 1996.) In biology courses we can always have students memorize a few more facts, and it is relatively easy for us to write exam questions that test students' ability to remember those facts. However, writing challenging questions (e.g., those that are higher on Bloom's taxonomic hierarchy) is sometimes difficult. A joke among teachers is that many of our multiple-choice questions are as easy as:

What color is Johnny's red wagon?

A. Red

B. 14

C. Please pass the sweet and sour shrimp.

D. A military solution

The above question is obviously a knowledge-based question in that it involves the recall of a fact, which just happens to be located in the question. Exams that include undemanding knowledge-based questions provide easy points for students and some testing experts will claim that such questions are useless because they do not discriminate between high- and low-performing students; virtually every student who studies the material (or, in the example above, reads the question) will get those questions right. Despite this valid criticism, many teachers use easy multiple-choice questions to reward studying and to help students build confidence in their understanding ("I know I got at least one right!"). However, a key to a good test is the presence of questions that discriminate between students who know and understand the material from those who don't; that is, questions that reward vigorous study and understanding. To help with this task, we have developed a new type of multiple-choice question.

Background to the New Question

Multiple-choice questions require students to make decisions based on their understanding of some content. If a student's understanding is accurate, we hope they can answer the questions correctly. However, teachers also hope that students who posses inaccurate understanding will answer those same questions incorrectly (i.e., the students will not be able to make the proper decisions). True-or-false questions typically require one decision per question. For example:

During the depolarization phase of an action potential, sodium gates are closed.

True or False.

Adding more distracters increases the number of decisions made by a student:

Which of the following gates are open during the depolarization phase of an action potential?

A. Sodium

B. Potassium

C. Calcium

D. Acetylcholine

The above question requires a student to make more decisions. For students who are guessing, there are four decisions, but for students who are somewhat prepared there are only two logical answers (sodium and potassium), and for those who are well prepared the question is as easy as "What color is Johnny's little red wagon?" For many college biology courses, this question would not discriminate between higher-performing and lower-performing students (both groups would answer the question correctly) because most teachers overtly state this concept in class. Hopefully, it is also well documented in students' notebooks and textbook.

The difficulty of a question increases as the number of correct options increases. For example:

How many of the following statements are true of action potential physiology?

- Depolarization requires the movement of sodium ions.

- Repolarization requires the movement of sodium ions.
- Action potentials require the movement of ions across membranes.
- The sodium-potassium pump is required to hyperpolarize a neuron.
- The sodium-potassium pump helps re-establish membrane gradients after an action potential has passed.

A. Zero (0) statements are true.

B. One statement is true.

C. Two statements are true.

D. Three statements are true.

E. Four or more statements are true.

The above question can be reduced to five different true-and-false (knowledge level) questions, but at issue is the number of decisions a student makes per question. In this one question, a student must make at least five decisions. In a quiz that covers only a small amount of material, the above question should probably be transformed into five different true-and-false questions, but for large exams that involve a limited number of questions per concept, the above one question can effectively discriminate between students with a good vs. excellent understanding of action potentials.

Our new multiple-choice question requires students to make several decisions per question—it requires them to *evaluate* the quality of a scientific explanation. We call these questions "You are the Teacher" questions because the students are role-playing a teacher correcting a short-answer exam (i.e., a form of reciprocal teaching). Reciprocal teaching is a broad educational idea that has much of its research foundation in reading education (Rosenshine & Meister, 1994) and is founded on the notion that students can learn when they take the role of teacher. At the center of reciprocal teaching is the notion that teachers are frequently evaluating work, and evaluation is at the highest level of Bloom's taxonomy. It is logical, then, that if students take the role of a teacher and are evaluating work, they then would be achieving the higher levels of Bloom's taxonomy. Teachers have for years had students exchange papers for grading, but this practice has typically not been part of the testing/learning process, but rather as a mechanism to reduce the work load of the teacher. "You are the Teacher" questions, however, require students to evaluate statements, similar to a teacher evaluating a short-answer essay question.

The following are the directions given for "You are the Teacher" questions, as well as two example questions:

Pretend you are a very nice science teacher who is correcting the following essays. How many scientific errors can you find in the following paragraph? Note: There is a maximum of one error per sentence.

Question 1

During the depolarization phase of an action potential, sodium gates are open and sodium diffuses from the extra-cellular fluid to the intra-cellular fluid. At the end of the depolarization phase, sodium gates close, and potassium gates open. Repolarization begins when potassium moves by active transport from the intra-cellular fluid to the extra-cellular fluid of the cell. After the action potential passes, ion gradients are maintained by the sodium/potassium pump.

A. 0 errors

B. 1 error

C. 2 errors

D. 3 errors

E. 4 or more errors

Question 2

An action potential reaches an axon terminal and calcium ions diffuse into the terminal. Synaptic vesicles fuse with the plasma membrane of the axon terminal and neurotransmitter substances are released into the synaptic cleft. The neurotransmitter substances move by active transport across the synaptic cleft and bind with receptors on the post synaptic membrane. The neurotransmitter/receptor complex contributes to the formation of another action potential on the post-synaptic neuron.

A. 0 errors

B. 1 error

C. 2 errors

D. 3 errors

E. 4 or more errors

The directions for the "You are the Teacher" questions were developed after a few trials in which students were including unintended grammatical errors; as a result we added the phrases "very nice" and "scientific errors" to clarify the requirements. We also deduced that there should be a maximum of one error per sentence because multiple errors in one sentence were difficult to identify. (For example, how many errors are in this sentence: "President Bush is a ballet dancer who is a quarterback for a neuron." Two? Three? Four?) The paragraphs evaluated by students are scientific concepts put into narratives; there is a scientific story line within each paragraph. For students who know and understand the concepts, errors should be easily identifiable. But for students with less accurate understandings, the errors—and more specifically the number of errors—are much more difficult to identify. This type of multiple-choice question requires considerable student *knowledge*, but moves to higher levels of Bloom's taxonomy because it

involves the *analysis* and *evaluation* of an explanation. This type of question can place a higher cognitive demand on a student than a basic knowledge question, but it can still be used within a multiple-choice exam.

Evaluation

A study examining the "You are the Teacher" questions' ability to discriminate between A, B, C, D and F-level students was performed in the fall of 2002 using students in a one-semester, freshman-level anatomy and physiology course taught in General College at the University of Minnesota. The questions were administered in two different settings. First, students practiced answering "You are the Teacher" questions during five quizzes given at different times throughout the semester. In this setting, students first answered the questions on an individual basis, and then also used the same questions for a group activity. The group activity was designed to help students understand the intent of the questions; they were able to watch and query other students while they proceeded through the questions. The second setting in which we administered the "You are the Teacher" questions was within large exams. The course had two 50-question exams (Tests 1 and 3), and two 100-question exams (Tests 2 and 4). Three "You are the Teacher" questions were used on the 50-question exams, and six were used within the 100-question exams.

Eighty-four students completed all the exams and were included in the evaluation. The following procedures were followed for the statistical analysis. First, all of the "You are the Teacher" points were summed across all four exams for each student (a maximum of 18 points). Next, the total points scored on the "You are the Teacher" questions were correlated with each student's total points earned in the course (a maximum of 500 total points) minus their points scored on the "You are the Teacher" questions. The "You are the Teacher" points were subtracted from the total points to make the former independent from the latter. We used total points because it represents a student's overall performance on exams, quizzes, lab reports, lab tests, etc.

A correlation of 0.81 was calculated for the "You are the Teacher" questions vs. total course points minus the "You are the Teacher" points (see Figure 1). Though far from a perfect correlation (i.e., 1.0), the data show a strong relationship between student performance on the "You are the Teacher" questions and students' overall performance in the course. These data indicate that the "You are the Teacher" questions are effective at discriminating between A, B, C, D, and F-level students.

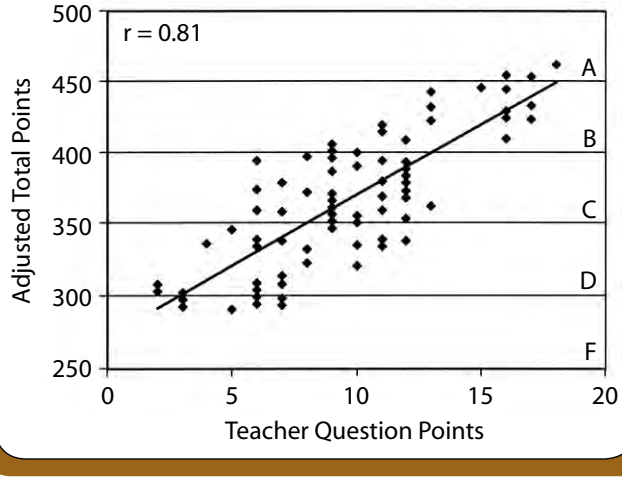
Conclusions

Over the 15 weeks of the semester, topics for the "You are the Teacher" questions included protein synthesis, mitosis and meiosis, muscle contraction, blood pressure regulation, nephron physiology, and several others. Additionally, questions that were initially used on quizzes were frequently modified and used on exams. Modifications normally involved altering key terms (e.g., changing diffusion to active transport, or sodium to potassium). All questions used in this study related to human anatomy and physiology, but they could be used in any course in which concepts can be written into story form.

During the semester students had ample opportunity to practice the "You are the Teacher" questions (e.g., individual and group quizzes and homework assignments), yet most students' reactions to the questions were negative. "Are you going to put those find the error questions on the next test, too?!" and "Could we please not have any more of those types of questions!" were typical comments. No student requested more "You are the Teacher" questions on exams, and many students wanted the questions to be eliminated. Despite these comments and opinions, virtually all students agreed that the questions required students to have a thorough understanding of the concepts, and that the questions were fair.

A single test comprised of only "You are the Teacher" questions would definitely be challenging for entry-level students, and their continued use would likely lead to high drop-out rates and low teacher evaluations. The mix of three questions out of 50 (Tests 1 and 3) and six questions out of 100 (Tests 2 and 4) seemed about right, and maybe a bit heavy. However, the majority of questions on those exams were knowledge-based and, as usual, many students answered those questions correctly. A logical next step in the development of the "You are the Teacher" ques-

Figure 1.
Scatter plot shows that "You are the Teacher" questions correlate well with overall grade achievement and discriminate well between high-end achievers (A and B students) and other students.
Teacher Question Points represent the total number of points earned on the "You are the Teacher" questions.
Adjusted Total represents the total number of points earned in the class minus the total number of points earned on the "You are the "Teacher" questions.



tions would be to evaluate their use with students in more advanced courses.

When placing a test question into a difficulty hierarchy, such as Bloom's taxonomy, it is necessary to keep your students' abilities in mind. A very difficult "You are the Teacher" question for our students (freshman) would likely be at the "evaluation" or "analysis" levels, but that same question used with more advanced students would probably fall into the "knowledge" level. In fact, an argument could be made that "You are the Teacher" questions are strings of knowledge-based questions that simply require "more knowledge." Much more in-depth research would be needed to truly say that our students are achieving the upper levels of Blooms taxonomy. However, what we can say with confidence is that students' performance on the "You are the Teacher" questions correlated very strongly with overall course performance in this specific course.

Biology courses, and especially anatomy and physiology courses, have long had the reputation of being "fact-based:" Students simply memorize a bunch of facts and then regurgitate them on exams that emphasize the knowledge level of Bloom's taxonomy. There will always be "facts" in science classes, and thus there will always be a place for knowledge-based questions on exams. However, it is important that science teachers stress a balance of cognitive tasks within our courses. Writing multiple-choice questions that achieve the upper levels of Bloom's taxonomy (analysis, synthesis, and evaluation) is difficult, but can be achieved through use of the "You are the Teacher" questions.

References

American Association for the Advancement of Science. (1993). *Benchmarks for Science Literacy*. New York: Oxford University Press.

Bloom, B. (Editor). (1956). *Taxonomy of Educational Objectives: Handbook 1, Cognitive Domain*. New York: David Mc Kay.

National Research Council. (1996). *National Science Education Standards.* Washington, DC: National Academy Press.

Moore, R., Jensen, M., Hsu, L. & Hatch, J. (2002). Saving the "false negatives:" Intelligence tests, the SAT, and developmental education. In D.B. Lundell & J.L. Higbee (Editors), *Exploring Urban Literacy & Developmental Education* (pp. 47-57). Minneapolis, MN: Center for Research on Developmental Education and Urban Literacy, General College, University of Minnesota.

Neill, A.S. (1960). *Summerhill: A Radical Approach to Child Rearing.* New York, NY: Hart.

Rosenshine, B. & Meister, C. (1994). Reciprocal teaching: A review of the research. *Review of Educational Research, 64*(4), pp. 479-530.

Smith, P.L. & Ragan, T.J. (1999). *Instructional Design.* New Jersey: Prentice-Hall, Inc.