

Mrs. Poindexter* was teaching Human Anatomy for the first time, and more than 120 students had signed up for the class! She worked long and hard to develop a midterm test to assess her teaching. She hoped that the 40 items on the multiple-choice test would provide some quick insights into how well her students understood the concepts. But when she returned the tests to the students and began to discuss the “right” answers, she was disturbed at how many questions the students raised. A number of students had interpreted some of the questions quite differently than she had intended. And some of her questions actually rewarded the students who didn’t understand the concepts instead of those who did!

Despite the well-documented drawbacks of multiple-choice questions, teachers and researchers use them for everything from kindergarten surveys to law school admission tests. Mrs. Poindexter chose a multiple-choice format for a number of reasons:

First, by including a large number of questions, she could evaluate her effectiveness in explaining key concepts. She realized that an objective, summative exam, by itself, cannot adequately measure student learning. But when carefully designed and combined with other evaluation measures, such as performance assessment, multiple-choice tests can play a legitimate role in assessment.

Developing High-Quality Multiple-Choice Test Questions

Second, a multiple-choice test appeared to be easy to create and grade, even for such a large class.

After asking an assessment specialist for advice, Mrs. Poindexter realized that while her choice of type of test was appropriate, many of her questions violated the guidelines for clarity and evaluated only lower-level thinking skills, such as recall and recognition. In addition, many of the questions provided inadvertent clues to the correct answers, which further decreased the test’s effectiveness in evaluating student understanding.

This article offers some suggestions for improving multiple-choice tests by decreasing “give-away” answers and developing questions that assess students’ higher-level thinking.

Despite the well-documented drawbacks of multiple-choice questions, teachers and researchers use them for everything from kindergarten surveys to law school admission tests.

BY GAIL TAYLOR RICE

Writing Multiple-Choice Questions

Teachers are expected to be able to write good questions for quizzes and tests—often with very little guidance or training. Because they rarely take formal courses on the art of test development, teachers often progress through several levels of expertise through trial and error.

For example, Mr. Seifert* has as one learning objective for his class in emergency medical care: *Students will be able to recognize common symptoms of internal hemorrhage.* Here is a quote from the textbook: “A patient experiencing hemorrhage, whether external or internal, might be expected to have a fast, thready pulse, cold, clammy skin, falling blood pressure, and a general state of agitation.”

* Not their real names.



Figure 1

Suggestions for Well-Designed Multiple-Choice Questions

- Design each item to measure an important learning objective.
- In the directions, tell students to select the “best” answer, rather than the “right” answer.
- Present all of the necessary information to answer the question in a clear, well-formulated stem.
- Make answers and distractors as short as possible (include any repeated information in the stem).
- Avoid negative wording in the stem (i.e., Which one of the following is not included?).
- Make distractors plausible and attractive to the uninformed.
- Be careful when using the words *always*, *never*, *all*, or *none*.
- Make the question (stem) concise and clear.
- Avoid complicated sentences.
- Include questions that measure higher cognitive functioning.

The first question Mr. Seifert developed looked like this:[†]
 1. A patient who is hemorrhaging might exhibit all of the following except:

- a. fast pulse
- b. cold, clammy skin
- c. falling blood pressure
- d. **feeling of well-being**

This question is very easy to create because you need only select three items from the list in the textbook and think of one wrong item to add. This question is typical of those developed by relatively inexperienced teachers.

After using the question for a while, Mr. Seifert realized that it would be better to ask what is true rather than what is false,¹ so he developed version 2 below:

2. Which of the following is the most likely sign of hemorrhaging?

- a. fast, thready pulse
- b. **warm, dry skin**
- c. hypertension
- d. sleepiness

This type of question is more challenging to write because the test developer includes only one answer from the textbook and must supply three incorrect but plausible responses, called distractors.

Here’s a third and better version of the question that uses the entire list, but in varying formats:

[†] Answers are in bold type.

Figure 2
Clued Quiz

Instructions: There is at least one clue in each of the following questions. Try to figure out which is the “right” answer. After circling the letter of your choice, consult Figure 3’s list of common clues. Put the number of the clue each question includes in the right-hand column. Answers are at the end of the article.

	Clue No.
1. The primary purpose of the stam is to remove the a. carm b. denton c. menice d. stambar	_____
2. Which of the following is/are most closely associated with the agreement? a. Stephen Douglass b. Robert E. Lee c. Abraham Lincoln d. All of the above	_____
3. Which of the following pairs has won the greatest number of Abby awards? a. Jones and Smith b. Smith and Taylor c. Smith and White d. White and Allen	_____
4. How many pounds of pressure is exerted by a Callam? a. 2.6 b. 150 c. 260 d. 2600	_____
5. The stanon is aided by a a. anstel b. immon c. octal d. port	_____
6. The stanon frequently overheats because a. all anstels are belious b. no immon is directly fectitious c. ports are always actial d. the octal is usually casable	_____
7. Stamation normally occurs when the a. anstels rupture b. immon falls and the denton is in place c. octal rotates easily d. ports pass over the carm	_____

Teachers are expected to be able to write good questions for quizzes and tests—often with very little guidance or training.

3. A patient who is hemorrhaging would exhibit
- a. warm, dry skin; hypotension; and a bounding pulse.
 - b. hypertension; a bounding pulse; and cold, clammy skin.
 - c. a weak, thready pulse; hypertension; and warm, dry skin.
 - d. **hypertension; cold, clammy skin; and a weak, thready pulse.**

This question shows a good way to test content derived from a list of conditions. A warning—when writing this kind of question, be careful not to overuse the correct options, thus giving a convergence clue. (See Figure 3.)

An even better way to test this kind of content would be to develop a case-based question, which requires application and higher levels of cognitive functioning. See Questions 4 and 5 below.

Figure 1 provides some guidelines for developing good multiple-choice questions.

Taking Clues Out of Multiple-Choice Questions

Occasionally, savvy students can guess the right answer to a question by looking for embedded clues. Figure 2 contains a simple quiz to help you check your own “testwiseness.” Figure 3 lists common types of clues teachers include in multiple-choice questions. See if you can figure out which type of clue is present in each of the questions in the quiz. You will find the answers at the end of the article.

By checking for inadvertent clues, teachers can make sure that their tests measure student mastery of content, rather than their test-taking savvy. Figure 4 provides a checklist to help you recognize and remove inadvertent clues from multiple-choice questions.

Writing Questions That Require Higher Cognitive Reasoning

Developing multiple-choice questions that require high-level thinking requires a great deal of thought and time. Professional test writers say that it can take an entire day to create more than three or four items, so don't wait until the night before a test to begin assembling the questions. Spread out the work over the quarter or semester, so you have good questions to draw upon to analyze the ongoing effectiveness of the teaching as well as questions to use on the final exam.

Here are some ideas to help you increase your supply of high-quality questions.

- Write a question that requires students to predict the outcome of a situation rather than simply label a phenomenon.
- Provide an abstraction or principle, and ask students to select an example that best illustrates the principle.
- Give examples, and ask for the principle or theory they illustrate.
- Create case studies, and base questions on them. Often, you can ask several questions based on each case and thereby expand the possibilities for assessing advanced levels of thinking. Here is how one teacher used several questions about one case study to enhance student understanding and learning:

4. A 14-year-old boy is brought to the emergency room following trauma experienced while playing baseball at school. He appears restless and agitated, although he does not seem to be in a great deal of pain. His pulse is 130/minute, his blood pressure is 100/50, and he has cold, clammy skin. Which of the following possibilities would be most important to assess immediately?

- broken bones
- internal bleeding**
- appendicitis
- psychiatric sequelae

5. The emergency room personnel send the boy to X-ray, and while waiting for the results, recheck his vital signs. The pulse is now 148/min and the blood pressure is 95/48. Which of the following would be the most important first action to take?

- Request a surgical consult re-**

garding possible internal bleeding.

- Call orthopedics for a possible bone setting, pending X-ray results.
- Ask a technician to draw blood to check for possible appendicitis.
- Call a family member to take the boy home.

Questions 4 and 5 take the objective to a new level. Now the teacher is asking for application and analysis, rather than

just knowledge, recognition, and recall. Students must know the significance as well as the definition of a fast pulse and high blood pressure. The questions assess whether they understand the cause of the symptoms and what to do about them.

- Provide a continuum of possibilities in your answers. Multiple-choice questions can also ask students to make predictions based on their interpretation of

Figure 3

Clues to Answers for Multiple-Choice Questions

1. Length Clue

The longest answer (or the one with the most qualifiers) is often the correct one because the writer has excessively qualified the information to ensure that it is accurate.

2. Verbal Association Clue

Sometimes questions suggest a verbal association between the stem and the correct answer. Watch for a different form of a key word, if it appears in only one answer.

3. Grammatical Clue

Each answer, with the stem, should form a grammatically correct statement. If not, this provides a clue to the right answer. For example, if the stem asks for a singular response, and all of the incorrect answers are plural, the student could correctly guess the right answer.

4. Specific Qualifiers

Qualifying words, such as *always*, *all*, or *never*, rarely appear in a correct answer, but are often found in distracters. On the other hand, less extreme qualifiers such as *usually*, *rarely*, *sometimes*, *typically*, and *may be* commonly appear in the correct option.

5. Convergence

Convergence refers to the tendency of test designers to use the correct answer more frequently than the incorrect options. Thus, the student can guess the best answer simply by counting which facts occur most frequently in the various answers, and selecting that response.

6. All of the Above / None of the Above

Research has shown that "all of the above" is the correct response 50-75 percent of the time, compared to "none of the above," which is the right choice less than 10 percent of the time.

7. Middle Ground

Particularly when using numbers as answers, test designers tend to use figures that are higher and lower when developing distracters. If you don't know an answer and wish to guess wisely, eliminate the lowest and highest numbers and choose one in the middle.

By checking for inadvertent clues, teachers can make sure that their tests measure student mastery of content, rather than their test-taking savvy.

quoted materials. Here are examples of questions that measure understanding and require some analysis of the following quotation from Walvoord and Anderson's *Effective Grading: A Tool for Learning and Assessment*.

"The social meaning of grading is chang-

Figure 4

Checklist for Eliminating Clues in Multiple-Choice Questions

- Avoid "all of the above" or "none of the above."
- Don't always place the right answer in the middle of a sequence (sometimes make all of the distractors larger or smaller than the right answer).
- Double check for grammatical clues—make sure all distractors fit grammatically with the stem.
- Try to make your distractors about the same length as the correct answer and of similar complexity.
- Check for verbal clues. If you repeat a word from the stem in the correct answer, make sure that word appears in some of the distractors, as well.
- Check to see if there is a pattern to the answers (e.g., A, C, D, A, C, D, B, D, A, B, D, A) or too many answers are in the same position (in a 10-point test: 6 B's, 2 A's, 1 C, and 1 D, for instance), and if so, rearrange the responses in several of the questions.



ing all the time. Your grades and grading system will be interpreted and used within the system that is—not the one you wish for or the one you experienced as a student. . . . Because grades are useless apart from the meanings that people impart to them, we suggest you abide by the system of meanings in which you find yourself. Except when issues of integrity and ethics are at stake, it's okay to pay for a hamburger at the current inflated rates, not the rates you paid in the good old days. Grade inflation is a national problem and must be addressed by institutions in concert at the national level. Individual teachers cannot address the problem in isolation; all you can do is use the coin of the realm."²

Question 1

When establishing a grading scale for a particular course, how should one determine which letter grade represents average work, according to Walvoord and Anderson?

- Stick to historical standards.
- Stick with established institutional norms.
- Use present national standards.**
- Develop different scales for each course, based on external standards.

Question 2

What do the authors mean when they suggest that in testing and grading, one should use the "coin of the realm"?

- Test using only the published course objectives.
- Base grades on national averages.**
- Ensure that grades represent student learning.
- Insist that grading be totally objective.

These questions illustrate attempts to develop multiple-choice questions that measure students' abilities to go beyond simply recalling facts, lists, or information. Remind students to read carefully, keeping in mind that although more than one answer may be true, they should choose only the response that can be deduced from the content of the paragraph.

Summary

Mrs. Poindexter tried not to become too discouraged with her early attempts to assess student knowledge through multiple-choice tests. She realized that

Developing multiple-choice questions that require high-level thinking requires a great deal of thought and time.

testing had a motivating effect on her students, as they studied to learn what they anticipated what would be on the test. Certainly, tests can aid retention and transfer of learning. And teachers need the feedback provided by students' test scores to help them evaluate their effectiveness. For tests to be valid, however, they must be carefully planned to accurately assess how well students have met learning objectives. By eliminating clues and designing questions that test high cognitive functioning, Mrs. Poindexter found she could increase the effectiveness of the testing process and thereby improve teaching and learning. ✍

This article has been peer reviewed.



Dr. Gail Rice is on the faculty of the School of Allied Health Professions at Loma Linda University in Loma Linda, California. As a faculty member in the new health professions higher education program, she recently

completed a Fellowship in Medical Education at the University of Southern California. She frequently presents workshops and courses on the topic of authentic assessment and consults with the University of Washington in this area. Dr. Rice authored an article on online assessment in the April/May 2003 issue of the JOURNAL.

Answers to Figure 2 quiz

For each of the quiz questions in Figure 1, the correct "clued" answer is given, followed by the number of the type of clue (from Figure 3) the question represents.

- d, Clue 2
- d, Clue 6

- c, Clue 5
- c, Clues 5 and 7
- d, Clue 3
- d, Clue 4
- b, Clue 1

SELECTED BIBLIOGRAPHY

- Cashin, William, *Improving Multiple-Choice Tests*. Manhattan: Center for Faculty Evaluation and Development in Higher Education, Kansas State University, 1986. Idea Paper No. 16.
- Davis, Barbara Gross, *Tools for Teaching*. San Francisco: John Wiley & Sons, Inc., 2001.
- Gronlund, Norman, *How to Make Achievement Tests and Assessments*. Boston: Allyn & Bacon, 1993.
- Jacobs, Lucy Cheser, and Clinton I. Chase, *Developing and Using Tests Effectively: A Guide for Faculty*. San Francisco: Jossey-Bass, 1992.
- Lowman, James, *Mastering the Techniques of Teaching*. San Francisco: Jossey-Bass, 1984.
- Maguire, Thomas, Ernest Skakun, and Charles Harley, "Setting Standards for Multiple-Choice Items in Clinical Reasoning," *Evaluation & the Health Professions* 15:4 (December 1992), pp. 434-452.
- Ory, John, *Improving Your Test Questions*. Urbana: Office of Instructional Resources, University of Illinois, 1985.
- Paxton, Moragh, "A Linguistic Perspective on Multiple-Choice Questioning," *Assessment & Evaluation in Higher Education* 25:2 (June 2000), p. 109.
- Quam, Kay J., *Ready, Set, Teach*. New York: Kay J. Quam, 1998.
- Schuwirth, L., C. van der Vleuten, and H. Jan Donkers, "A Closer Look at Cueing Effects in Multiple-Choice Questions," *Medical Education* 30:1 (1996), pp. 44-49.
- Skakun, E., T. Maguire, and D. Cook, "Strategy Choices in Multiple-Choice Items," *Academic Medicine* 69:10, Suppl. (October 1994), pp. S7-S9.
- Vitale, Barbara A., and Patricia M. Nugent, *TEST SUCCESS: Test Taking Techniques for the Healthcare Student*. Philadelphia: F. A. Davis Company, 1996.
- Walvoord, Barbara E., and Virginia Johnson, *Effective Grading: A Tool for Learning and Assessment*. San Francisco: Jossey-Bass, 1998.

NOTES AND REFERENCES

- In fact, advanced test developers almost always avoid using a negative question, e.g., Choose the exception or the one that is **not** correct.
- Barbara E. Walvoord and Virginia Johnson Anderson, *Effective Grading: A Tool for Learning and Assessment* (San Francisco: Jossey-Bass, 1998), p. 12.