

for your consideration...

suggestions and reflections on Teaching and Learning

CTL Number 8

November 1990

Improving Multiple Choice Questions

Although multiple choice questions are widely used in higher education, they have a mixed reputation among faculty because they seem inherently inferior to essay questions. Many teachers believe that multiple choice exams are only suitable for testing factual information, whereas essay exams test higher-order cognitive skills. Poorly-written multiple choice items elicit complaints from students that the questions are “picky” or confusing. But multiple choice exams *can* test many of the same cognitive skills that essay tests do, and, if the teacher is willing to follow the special requirements for writing multiple choice questions, the tests will be reliable and valid measures of learning. In this article, we have summarized some of the best information available about multiple choice exams from the practices of effective teachers at UNC and from testing and measurement literature.

Validity and Reliability

The two most important characteristics of an achievement test are its content validity and reliability. A test's validity is determined by how well it samples the range of knowledge, skills, and abilities that students were supposed to acquire in the period covered by the exam. Test reliability depends upon grading consistency and discrimination between students of differing performance levels. Well-designed multiple choice tests are generally more valid and reliable than essay tests because (a) they sample material more broadly; (b) discrimination between performance levels is easier to determine; and (c) scoring consistency is virtually guaranteed.

The validity of multiple choice tests depends upon systematic selection of items with regard to both content and level of learning. Although most teachers try to select items that sample the range of content covered by an exam, they often fail to consider the level of the questions they select. Moreover, since it is easy to develop items that require only recognition or recall

of information, they tend to rely heavily on this type of question. Multiple-choice tests in the teacher's manuals that accompany textbooks are often composed exclusively of recognition or recall items.

Cognitive Levels

Teachers are aware of the fact that there are different levels of learning, but most of us lack a convenient scheme for using the concept in course planning or testing. Psychologists have elaborate systems for classifying cognitive levels, but a simple three-level scheme is sufficient for most test planning purposes.

We will designate the three categories *recall*, *application* and *evaluation*. At the lowest level, *recall*, students remember specific facts, terminology, principles, or theories—e.g., recollecting the equation for Boyle's Law or the characteristics of Impressionist painting. At the median level, *application* students use their knowledge to solve a problem or analyze a situation—e.g., solving a chemistry problem using the equation for Boyle's Law or applying ecological formulas to determine the outcome of a predator-prey relationship. The highest level, *evaluation*, requires students to derive hypotheses from data or exercise informed judgment—e.g., choosing an appropriate experiment to test a research hypothesis or determining which economic policy a country should pursue to reduce inflation. Since it is difficult to construct multiple choice questions at the complex application level, we have included an example at the end of this article.

Analyzing your course material in terms of these three categories will help you design exams that sample both the range of content and the various cognitive levels at which the students must operate. Performing this analysis is an essential step in designing multiple choice tests that have high validity and reliability. The test matrix, sometimes called a “table of specifications,” is a convenient tool for accomplishing this task.

The Test Matrix

The test matrix is a simple two-dimensional table that reflects the *importance* of the concepts and the *emphasis* they were given in the course. At the initial stage of test planning, you can use the matrix to determine the proportion of questions that you need in each cell of the table. For example, the chart on the right reflects an instructional unit in which the students spent 20% of their time and effort at the recall and simple application levels on Topic I. On the other hand, they spent 25% of their effort on the complex application level of Topic II. This matrix has been simplified to illustrate the principle of proportional allocation; in practice, the table will be more complex. The greater the detail and precision of the matrix, the greater the validity of the exam.

As you develop questions for each section of the test, record the question numbers in the cells of the matrix. The table below shows the distribution of 33 questions on a test covering a unit in a psychology course.

Topics	Cognitive Levels		
	Recall	Application	Evaluation
I	10%	10%	5%
II	5%	10%	25%
III	5%	10%	20%

In practice, you may have to leave some of the cells blank because of limitations in the number of questions you can ask in a test period. Obviously, these elements of the course should be less significant than those you do choose to test.

Topics	Recall	Application	Evaluation	
A. Identify crisis vs. role confusion; achievement motivation.	2, 9	4, 21, 33	16	18%
B. Adolescent sexual behavior; transition of puberty.	5, 8	1, 13, 26	11	18%
C. Social isolation and self esteem; person perception.	14, 6	3, 20	25	15%
D. Egocentrism; adolescent idealism.	7, 29	12, 31	10, 15, 27	21%
E. Law and maintenance of the social order.	17	22	18	9%
F. Authoritarian bias; moral development.	19	30	24	9%
G. Universal ethical principle orientation.	28	23	32	9%
	33%	40%	27%	

Guidelines for Writing Questions

Constructing good multiple choice items requires plenty of time for writing, review, and revision. If you write a few questions after class each day when the material is fresh in your mind, the exam is more likely to reflect your teaching emphases than if you wait to write them all later. Writing questions on three-by-five index cards or in a word-processing program will allow you to rearrange, add, or discard questions easily.

The underlying principle in constructing good multiple choice questions is simple: the questions must be asked in a way that neither rewards "test wise" students nor penalizes students whose test-taking skills are less developed. The following guidelines will help you develop questions that measure learning rather than skill in taking tests.

Writing the Stem

The “stem” of a multiple-choice item poses a problem or states a question. The basic rule for stem-writing is that students should be able to understand the question without reading it several times and without having to read all the options.

1. Write the stem as a single, clearly-stated problem. Direct questions are best, but incomplete statements are sometimes necessary to avoid awkward phrasing or convoluted language.
2. State the question as briefly as possible, avoiding wordiness and undue complexity. In higher-level questions the stem will normally be longer than in lower-level questions, but you should still be brief.
3. State the question in *positive* form because students often misread negatively phrased questions. If you must write a negative stem, emphasize the negative words with underlining or all capital letters. Do not use double negatives—e.g., “Which of these is not the least important characteristic of the Soviet economy?”

Writing the Responses

Multiple-choice questions usually have four or five options to make it difficult for students to guess the correct answer. The basic rules for writing responses are (a) students should be able to select the right response without having to sort out complexities that have nothing to do with knowing the correct answer and (b) they should not be able to guess the correct answer from the way the responses are written.

1. Write the correct answer immediately after writing the stem and make sure it is *unquestionably* correct. In the case of “best answer” responses, it should be the answer that authorities would agree is the best.
2. Write the incorrect options to match the correct response in length, complexity, phrasing, and style. You can increase the believability of the incorrect options by including extraneous information and by basing the distractors on logical fallacies or common errors, but avoid using terminology that is completely unfamiliar to students.
3. *Avoid* composing alternatives in which there are only microscopically fine distinctions between the answers, unless the ability to make these distinctions is a significant objective in the course.
4. *Avoid* using “all of the above” or “both A & B” as responses, since these options make it possible for students to guess the correct answer with only partial knowledge.
5. Use the option “none of the above” with extreme caution. It is *only* appropriate for exams in which there are absolutely correct answers, like math tests, and it should be the correct response about 25% of the time in four-option tests.

6. *Avoid* giving verbal clues that give away the correct answer. These include: grammatical or syntactical errors; key words that appear only in the stem and the correct response; stating correct options in textbook language and distractors in everyday language; using absolute terms—e.g., “always, never, all,” in the distractors; and using two responses that have the same meaning.

General Issues

1. All questions should stand on their own. *Avoid* using questions that depend on knowing the answers to other questions on the test. Also, check your exam to see if information given in some items provides clues to the answers on others.
2. Randomize the position of the correct responses. One author says placing responses in alphabetical order will usually do the job. For a more thorough randomization, use a table of random numbers.

Analyzing the Responses

After the test is given, it is important to perform a test-item analysis to determine the effectiveness of the questions. Most machine-scored test printouts include statistics for each question regarding item difficulty, item discrimination, and frequency of response for each option. This kind of analysis gives you the information you need to improve the validity and reliability of your questions.

Item difficulty is denoted by the percentage of students who answered the question correctly, and since the chance of guessing correctly is 25%, you should rewrite any item that falls below 30%. Testing authorities suggest that you should strive for items that yield a wide range of difficulty levels, with an average difficulty of about 50%. Also, if you expect a question to be particularly difficult or easy, results that vary widely from your expectation merit investigation.

To derive the item discrimination index, the class scores are divided into upper and lower halves and then compared for performance on each question. For an item to be a good discriminator, most of the upper group should get it right and most of the lower group should miss it. A score of .50 on the index reflects a maximal level of discrimination, and test experts suggest that you should reject items below .30 on this index. If equal numbers of students in each half answer the question correctly, or if more students in the lower half than in the upper half answer it correctly (a negative discrimination), the item should not be counted in the exam.

Finally, by examining the frequency of responses for the incorrect options under each question, you can determine if they are equally distracting. If no one chooses a particular option, you should rewrite it before using the question again.

Developing Complex Application Questions

Devising multiple choice questions that measure higher-level cognitive skills will enable you to test such skills in large classes without spending enormous amounts of time grading. In an evaluation question, a situation is described in a short paragraph and then a problem is posed as the stem of the question. All the rules for writing multiple choice items described above also apply to writing evaluation questions, but students must use judgment and critical thinking to answer them correctly. In the example below (adapted from Welsh, 1978), students must understand the concepts of price inflation, aggregate private demand, and tight monetary policy. They must also be able to analyze the information presented and, based on projected effects, choose the most appropriate policy. This question requires critical thinking and the complex application of economic principles learned in the course.

“Because of rapidly rising national defense expenditures, it is anticipated that Country A will experience a price inflation unless measures are taken to restrict the growth of aggregate private demand. Specifically, the government is considering either (1) increasing personal income tax rates or (2) introducing a very tight monetary policy.” If the government of Country A wishes to minimize the adverse effect of its anti-inflationary policies on economic growth, which one of the following policies should it use ?

- a. The tight money policy because it restricts consumption expenditures more than investment.
- b. The tight money policy, since the tax increase would restrict consumption expenditures.

- c. The personal income tax increase since it restricts consumption expenditures more than investment.
- d. Either the tight money policy or the personal income tax rate increase since both depress investment equally.

Teachers have developed similar questions for analysis and interpretation of poetry, literature, historical documents, and various kinds of scientific data. If you would like assistance in writing this kind of multiple choice question, or if you would like to discuss testing in general, please call the Center for Teaching and Learning for an appointment.

Bibliography

Cashin, W.E. (1987) *Improving multiple-choice tests* Idea Paper No. 16. Manhattan, KS: Center for Faculty Development and Evaluation, Kansas State University.

Gronlund, N.E. (1988) *How to construct achievement tests* (4th ed.). Englewood Cliffs, NJ: Prentice-Hall.

Sax, G. (1974) *Principles of educational measurement and evaluation*. Belmont, CA: Wadsworth.

Welsh, A.L. (1978) Multiple choice objective tests. In P. Saunders, A.L. Welsh, & W.L. Hansen (Eds.), *Resource manual for teacher training programs in Economics* (pp 191-228). New York: Joint Council on Economic Education.

