Teacher-Made Exams: Part 3

Gary A. Holt
Kay E. Holt

INTRODUCTION

The previous articles in this series have considered the philosophy of test writing and some general test construction guidelines, such as planning and organization. This discussion considers the types of test questions commonly used in test construction. This is an important consideration because an understanding of the advantages, disadvantages, and item-writing guidelines helps to ensure that tests are valid and reliable indicators of student knowledge. It is generally agreed among educational scholars that there are some rules governing the construction of each type of test item (1). Poorly written test items can actually discriminate against students with above-average knowledge and can discourage creative thinking. Test items that lack validity and reliability due to construction problems result in test scores that do not represent actual student knowledge. Finally, without an appreciation of the strengths, limitations, and guidelines for each item type, teachers are often tempted to place an undue emphasis on course content that is easily formulated into certain types of items, even when the content does not merit emphasis.

Gary A. Holt, M.Ed., Ph.D., is Assistant Professor of Pharmacy at the School of Pharmacy, University of Wyoming, P.O. Box 3375, University Station, Laramie, WY 82071-3375. Kay E. Holt is a doctoral student in the School of Education at the University of Wyoming.

TRUE/FALSE ITEMS

True/false items are possibly the most controversial of all testing approaches because their disadvantages tend to outweigh their advantages, yet they remain popular among many educators. They can be useful indicators of student knowledge if used judiciously and appropriately. It is important for educators to fully appreciate the strengths and weaknesses of true/false items and to take steps to maximize the beneficial aspects (2).

True/false items typically consist of a declarative statement that is either true or false as written. The task of the student is to determine which and answer accordingly. Table 1 summarizes the advantages of true/false items. Ironically, ease of construction is the most controversial advantage. Ostensibly, these questions would appear to be easy to construct, yet it can be argued that good true/false items are actually quite difficult to construct. This point will be considered in greater detail in the discussion on advantages and guidelines for writing.

Because students usually are only required to agree or disagree with the statement, true/false items are easily scored. Scoring is relatively mechanical, and a high degree of objectivity is possible (2-4). These items allow for a wide sampling of topics and course content because students can usually read and answer them quite quickly; therefore, many true/false items can be included on a single exam (1, 3, 4). They are amenable to many subject areas and situations and are considered to be a realistic task for students, since everyday life requires people to judge situations as being true or false (1, 4).

As indicated above, many educators consider the disadvantages of true/false items to far outweigh any benefits. And, it is likely that

TABLE 1. Advantages of True/False Questions

- Can be constructed easily and rapidly
- Can be scored easily
- Allow for wide sampling of topics and course content
- Can be used in many subject areas and situations
- Considered to be a realistic task for the pupil

this type of item is used far more often than it should be. The disadvantages for true/false items are summarized in Table 2. A particular concern of this type, from an educational perspective, is that it emphasizes memorization, which is a low-level cognitive skill. In part this is because it is quite difficult to construct good true/false items that involve generalization, broad principles, relationships, comprehension, applications, analysis, synthesis, and evaluation (1, 3). Modern educational philosophies deemphasize memorization, except as a means of gaining prerequisite knowledge for more complex skills. What students can do with a skill is more important than the fact that they have it (3).

True/false items are associated with a high guessing factor because students are confronted with only two alternatives for an answer. This means that they have a 50/50 chance of guessing the correct answer even when they do not know the material (1, 3, 4).

One of the difficulties in writing good true/false items involves their dependence upon absolute judgments. These statements usually presume situations that are either true or false, with no possibility for intermediate values. In truth, most facts are not entirely true or false, but require qualification. It is unfair to ask the student to guess at the teacher's criteria for evaluating the truth of a statement.

True/false statements usually measure only simple factual information. All too often, this concerns small, relatively unimportant pieces of information that are easily translated into true/false items (1, 4). It is worth mentioning that tests should be designed to reflect course content. Course content should not be manipulated to reflect convenient test designs.

Because of the disadvantages of true/false items, they are of limited help in diagnosing student, teacher, or course strengths and weaknesses. Ideally, tests should not merely be a reflection of student progress but should also clue teachers regarding teaching effectiveness and the appropriateness of course content. This can only occur when items are valid and reliable.

An interesting objection made to true/false items involves the concept of retroactive inhibition. If a student believes a statement to be true when in fact it is false, the test may reinforce the retention of misinformation. Later, it may be more difficult to unlearn the misinformation. Previous learning that interferes with present learning

TABLE 2. Disadvantages of True/False Questions

- Emphasize memorization
- Have high guessing factor
- Depend upon absolute judgments that may not actually exist
- Tend to measure only simple factual information
- Tend to be concerned with small, relatively unimportant, pieces of information easily formulated into true/false statements
- May be affected by extraneous factors, such as construction factors and key words
- Are of little help in diagnosing student, teacher, or course strengths and weaknesses
- Are actually difficult to write well
- Tend to be ambiguous, misleading students
- May result in retroactive inhibition, which can actually interfere with learning

is called retroactive inhibition, and this factor is seen by some educators as a serious argument against the use of true/false tests (2).

Table 3 presents guidelines for writing effective true/false items. First, teachers should avoid the use of specific determiners. A specific determiner is an unintentional clue to the correct answer. For example, the words "only," "all," "always," "no," "never," "every," and "none" are often associated with false statements. The words "usually," "sometimes," "often," "could," "generally," "customarily," "may," and "frequently" are often associated with true statements (1-4).

Every effort should be made to write items that are unequivocally true or false. Items should not trick students into giving incorrect answers because the objective is to test the students' knowledge and not their ability to take a test. Items should be written in a simple and direct manner and should be definite and unambiguous in meaning (2, 4). Underlining key terms and words that affect the meaning of the directions or items (e.g., *not*) helps to ensure that the students understand the information being presented (2). It is usually best to write each item in a positive rather than a negative form because negative statements (especially those containing double negatives) are more confusing (3, 4).

Items should measure important course objectives. All too often, insignificant facts and pieces of information are used only because they conveniently conform to a true/false format (3). Each item should be kept as short as possible and should focus on one basic idea. Complex sentences and those addressing multiple ideas are more confusing, especially when one part of the statement is true and another part is false (1, 2). It is best if all statements are of a relatively uniform length. Without planning, shorter statements are more likely to be false, while longer statements are more likely to be true. This occurs because more wording is usually required to qualify a true statement (2).

Teachers should avoid taking statements directly from a text and making only simple word changes. This practice only requires students to memorize, not to use higher-level cognitive skills. It is better to make significant wording changes. If students understand concepts, they should be able to recognize them regardless of the way in which they are stated. Varied wording challenges students to

TABLE 3. Guidelines for Writing True/False Items

- Avoid the use of specific determiners that may clue the student to the correct answer.
- Each statement should be unequivocally true or false or as absolutely true or false as possible.
- Items should not trick students into answering incorrectly.
- Items should be written as simply and directly as possible. They should be definite and unambiguous in meaning.
- Underline all words that affect meaning (e.g., <u>not</u>) and all key terms, both in the test directions and in individual test items.
- If possible, state each item in a positive rather than a negative form.
- Construct items that measure important objectives, avoiding insignificant facts and information.
- Keep items relatively short, focusing on only one basic idea.
- Try to write items of uniform length.
- Avoid taking statements directly from the textbook and making only simple word changes.
- Make approximately half of the statements false and half true.
- Make it easy for students to indicate their responses.

apply concepts to new situations. Also, when textbook wording is used, false statements usually sound false, even though the student may not know why they are false (2).

It is generally thought that approximately half of the items should require a true response and half a false response (2-4). Planning is important because true items are usually easier to construct than false ones. Thus, true items will occur more often (2, 3). In part, this happens because we think in terms of what is true rather than what is not true (2). Balancing the responses for true and false items is less distracting to students, who often look for patterns in responses (1, 4).

Finally, teachers should avoid using coding systems that may be confusing to students. For example, using a " + " to indicate true and a " − " to indicate false requires students to attend to yet another task while they are taking the exam (2). Validity is improved when testing instructions are kept simple.

Some variations of the traditional true/false format are occasionally used with success. These are summarized in Table 4. The yes/no, agree/disagree, and fact/opinion variations have the same advantages and limitations as the true/false format. Asking students to qualify their answers is possibly the best way to strengthen true/false and similar items because the qualification of answers demands greater cognitive skills on the part of the students and reduces the guessing element (4).

MULTIPLE-CHOICE ITEMS

The multiple-choice exam is currently the most popular format in use for both teacher-made and standardized tests (2). Multiple-choice items consist of two parts: a stem and answer options or alternatives. The stem is either a question or a statement that is to be

TABLE 4.   Variations on the Basic True/False Format

- Yes/No
- Agree/Disagree
- Fact/Opinion
- Qualification or correction

answered or completed by selecting one of the answer options. All of the incorrect or less appropriate options are called distracters. The student's task is to select the correct answer or best alternative from among the available options (1-4).

Table 5 lists the advantages of multiple-choice items. Like true/false items, they are quickly and easily scored (especially when machine-scorable answer sheets are used), allow for a large sampling of course content (students can read and answer them quickly), and can be written to cover the content of most subject areas and academic levels (1-4). Multiple-choice items reduce guessing because they usually have four or five answer options from which students can choose. Scoring is highly objective because little interpretation of student responses is required (1, 3, 4).

This type allows the teacher to design items for all cognitive levels (i.e., from memorization to the more complex levels of generalization, broad principles, relationships, comprehension, applications, analysis, synthesis, and evaluation) (1-4). And, it allows for the development of items that require students to discriminate among options that vary in degree of correctness. This can be identification of a "best" answer, which avoids the absolute judgments usually required for true/false items (3). Finally, multiple-choice items are particularly amenable to statistical analysis procedures that can identify areas of student weakness, ambiguities within the test, and teaching and course effectiveness (3, 4). This is a particular advantage for teachers who are striving to improve their educational efforts.

The disadvantages proposed for multiple-choice items are summarized in Table 6. The major problem is that writing good items can be quite difficult. Building the stems can be a demanding task, and identifying appropriate distracters can be hard. Thus, the development of good multiple-choice items is time-consuming for teachers (1-4). Another limitation is the time required for students to answer these items. Good reading comprehension is often a prerequisite for analysis and differentiation among answer options, so students require more time per question. This is especially true when the test requires students to demonstrate fine discriminations and fundamental understandings when selecting from among answer options (i.e., higher cognitive skills) (1, 4). It is worth mentioning

TABLE 5.   Advantages of Multiple-Choice Items

- Allow measuring of objectives for all cognitive levels, from memorization to the most complex levels
- Quick and easy to score, especially if machine-scorable answer sheets are used
- Allow a large sampling of course content due to small amount of time required to read and answer each question
- Wide adaptability in covering content in most subject fields
- Scoring highly objective
- Allow for items that require students to discriminate among options that vary in degree of correctness; allow students to select the best alternative and avoid absolute judgments usually required for true/false tests
- Reduce guessing
- Amenable to all grade or academic levels
- Amenable to statistical (item) analysis to detect areas of student weakness, item ambiguity, item difficulty, teaching and course effectiveness, and improvement of future tests

TABLE 6.   Disadvantages of Multiple-Choice Items

- Can be difficult and time consuming to write
- Require good reading comprehension for analysis and differentiation among options
- Require more time per question on the part of students

that the need for good reading comprehension is not in itself a limitation of multiple-choice items. It is simply that tests that demand higher-level cognitive skills will require more time to complete. Educators must plan accordingly so that students will have ample time to finish the test.

Table 7 provides guidelines for writing multiple-choice items. As with any test type, items should be written in clear and simple language (4). Teachers should use vocabulary suited to the maturity and academic level of the students but should avoid specific determiners and grammar clues that indicate the correct answer (2-4). Underline all words that affect meaning, both in the test directions and in the test items (2, 3). It is best to state stems and options in positive rather than in negative terms whenever possible, especially avoiding double negatives (3, 4). It is best to arrange items in order of increasing difficulty or according to topic because this helps students to organize their thinking (2). Items that measure opinions rather than facts should be avoided (3).

Stems should be as brief as possible. However, the stem should contain as much of the item as possible and should be written to avoid repeating words or phrases in the response options. It is better to have a long stem and short response options (1, 3, 4). Finally, the stem should be grammatically correct relative to the options. Grammatical problems often involve the use of ''a'' versus ''an'' or the use of inappropriate verb tenses. Most often, problems in grammatical consistency are due to carelessness on the part of instructors (1-4). Evidence does not suggest that stems written as questions are more or less effective than those written as statements. If the stem is open-ended, it should express a complete idea so that the student is certain about what is being asked (1, 3). Regardless of the form in which the stem is written, it should introduce what is expected of the student. It should address one central problem or question, which is clearly stated (3).

All distracters should be functional (i.e., plausible to the student). The teacher should avoid use of options that have similar or identical meanings or sound more correct than others (1-4). Usually it is best to use at least three options but no more than five. The length of the options should not be related to their tendency to be correct or incorrect (1, 2).

TABLE 7. Guidelines for Writing Multiple-Choice Items

- Write items in clear and simple language.
- Make the stem as brief as possible. However, it is better to have the stem long and the responses short. The stem should include as much of the item as possible.
- Make the stem clear and grammatically correct relative to the options.
- Include in the stem all words that would otherwise occur in each response option.
- Arrange items in order of increasing difficulty or according to topic.
- Make the stem introduce what is expected of the student and make the purpose of the test item clear. The stem should deal with only one central problem, which is clearly stated.
- Avoid specific determiners and grammar clues that indicate correct answers.
- Use vocabulary suited to the maturity and academic level of the students.
- State stems and options positively whenever possible. Avoid double negatives.

- Underline all words that affect meaning (e.g., <u>not</u>) and all key terms, both in the test directions and in individual items.
- Avoid items that measure opinions.
- Make all distracters functional (i.e., plausible).
- Vary the placement of the correct option. Each option position should have an equal chance to be the location of the correct answer.
- Vary the number of options as needed, using at least three and no more than five options.
- Be certain that the length of the options is not related to their tendency to be correct or incorrect.
- Avoid overlapping options (i.e., those that have similar or identical meanings).
- Avoid clang associations (i.e., items for which one option sounds more correct than other options).
- Use responses such as "none of the above" and "all of the above" with discretion. Avoid using them as fillers.
- Make each item have a correct or clearly best answer based on a comprehensive knowledge of the course content.
- Place response options in alphabetical, numerical, or some other logical order whenever possible.

Each item should have a correct option that is clearly the best alternative based upon a comprehensive knowledge of the course content (1-4). The location of the correct option should vary from item to item (2-4). There are a number of methods that teachers can use to determine the position of the correct answer without investing inordinate amounts of time trying to balance the answers. The simplest is to arrange the response options in alphabetical, numerical, or some other logical order and to allow the correct answer to fall where it may within this arrangement (4). It is also advisable to inform the students of this policy so that they will not spend time searching for answer patterns.

Finally, teachers should generally avoid use of "all of the above" and "none of the above" options. All too often these responses are used as fillers when the instructor is unable to write more desirable options. These options can be used effectively at times if the item is written carefully. However, these responses often serve as weak discriminators for testing purposes (2-4). A similar concern can be expressed about the use of "A and B of the above," "A and C of the above," and so forth. Teachers have sometimes managed to create 12 to 15 response options using this approach. This serves no true educational purpose and unnecessarily increases item response times because students are forced to sort out the bewildering array of options made available to them. There is rarely any useful purpose served by including more than five or six options on a multiple-choice item.

Table 8 lists methods by which multiple-choice items can be used to address higher cognitive skills. These are only intended to serve as representative examples of many possibilities.

SHORT-ANSWER ITEMS

Short-answer items are also referred to as completion items, fill-in-the-blank items, or supply items. Some writers make a minor distinction between short-answer and supply items versus completion and fill-in-the-blank items. The first two can require longer answers than the last two. In the present discussion no distinction will be made (1-3). All of these items require students to supply an answer to a question or statement rather than to select it from a

TABLE 8. Using Multiple-Choice Items to Test for Higher Cognitive Skills

- Construct items in a different form than that which was originally presented.
- Provide for a condition contrary to fact.
- Have students discover relationships among similar topics.
- Have students identify assumptions and analyze criteria.
- Have students select examples of principles or concepts.
- Use novel pictorial materials to measure principles that require students to apply knowledge.
- Use charts and tables.

group of alternatives. Usually the question or statement can be answered with a word, number, phrase, or sentence, but in all cases, it is a much abbreviated response relative to essay questions (1-3).

Table 9 summarizes the advantages of short-answer items. Construction is relatively easy, and scoring is quick and easy if the items are well written (1, 3, 4). Because each item can be quickly read and analyzed by students and because responses are brief, it is usually possible to sample a large volume of the course content using this type of item (3, 4). Guessing is low and reliability is high because the correct answer is not available for selection in the item (1, 3, 4). Perhaps the strongest advantage is that short-answer items measure recall more than recognition. They are especially useful for measuring student recall of names, dates, terms, places, definitions, and the like (2, 4).

The two specific disadvantages of short-answer items are summarized in Table 10. First, they generally only measure lower cognitive skills (e.g., memorization) instead of higher skills (e.g., understanding, application) (1, 3, 4). Conceivably, short-answer items could ask students to supply responses derived from higher-level cognitive skills. However, the short answer precludes the opportunity for students to relate the process by which they derived the answer. The second disadvantage involves scoring and interpretation difficulties that can arise from misspelled words, illegible writing, and evaluation of the correctness of synonyms and similar ideas. This subjectivity in scoring poses a significant problem for many teachers who attempt to use this type of item (1-4).

Table 11 gives guidelines for constructing appropriate short-answer items. The statement should be clearly written so that the correct answer is limited to one or two specific words. Only significant or key words should be omitted, and the information requested should be limited to one idea or topic. Asking for multiple pieces of information in the same statement creates confusion (1, 4). As with other test types, teachers should avoid specific determiners that clue the student to the correct answer. Items should clearly indicate the type of response requested (e.g., date, place, person, units of measurement). Otherwise, the student may be required to read the mind of the instructor, and the question ceases to be a valid indicator of student knowledge (1, 3, 4). Teachers should avoid taking state-

TABLE 9. Advantages of Short-Answer Items

- Construction relatively easy
- Especially good for measuring student recall of names, dates, terms, places, definitions, etc.
- Large sampling of course content possible due to small amount of time required to read and answer each question
- Low guessing factor
- Tend to measure recall more than recognition of information
- Reliability generally high due to a low guess factor and need to recall information
- Easily and quickly scored if written well

TABLE 10. Disadvantages of Short-Answer Items

- Generally only measure lower cognitive levels rather than higher levels
- Can be difficult to score due to subjectivity, misspelled words, illegible writing, and evaluation of correctness of synonyms and similar ideas

TABLE 11.   Guidelines for Writing Short-Answer Items

- Write a clear statement so that the correct answer is limited to one or
  two specific words.
- Omit only significant or key words.
- Limit the number of blanks, and make the number of blanks correspond to
  the answer.
- Make the omitted part of the statement occur toward the end of the
  sentence so that the question/statement is defined for the student prior
  to the time an answer is required.
- Avoid specific determiners that clue the student to the answer.
- Clearly indicate the type of response required (e.g., date, place,
  person).
- Avoid mutilating the statement until its meaning is all but lost.
- Ask for only one piece of information per statement or question.
- Avoid using statements taken directly from the textbook.

ments directly from the text because doing so only requires students to memorize (1, 3).

The primary problem in writing short-answer items is the number of blanks placed within the statement or question. The number of blanks should be limited and should correspond to the answer requested. Thus, if a two-word response is required, two blanks should be provided. Use of excessive numbers of blanks in one sentence mutilates the sentence to the point that its meaning is lost (1, 3, 4). Finally, the short-answer item should be written in such a way that the omitted portion occurs near the end of the sentence rather than at the beginning. Construction of the items in this way serves to define the question/statement for the student prior to the point at which an answer is required (1, 3, 4). In this way, the item is least likely to be confusing.

CONCLUSION

Educational scholars suggest that there is an art to writing tests (1). Test writing depends heavily upon the teacher's understanding of the course content and the students. It also depends heavily upon an understanding of the strengths and limitations of each item type, as well as adherence to appropriate guidelines for item construction. When teachers approach test writing as an art instead of a necessary task, their efforts are enhanced. Testing becomes a more accurate measure of student accomplishments and can even serve as a valuable learning tool for both teacher and students.

REFERENCES

1. Ahmann JS, Glock MD. Measuring and evaluating educational achievement. Boston: Allyn and Bacon, Inc., 1971.

2. Bergman J. Understanding educational measurement and evaluation. Boston: Houghton Mifflin Co., 1981.

3. Sax G. Principles of educational measurement and evaluation. Belmont, CA: Wadsworth Publishing Co., Inc., 1974.

4. Mitchell R, Westerman JE. Evaluation in the classroom. Dubuque, IA: Kendall/Hunt Publishing Co., 1977.