Teacher-Made Exams: Part 4

Gary A. Holt
Kay E. Holt

MATCHING TESTS

Generally speaking, matching items are a multiple-choice format in which the student associates an item in one column with a choice in a second column. In this typical format, the left column consists of the premise statements and the right column, the responses. A perfect matching situation exists when the number of premises is equal to the number of responses and the responses can only be used once. Imperfect matching exists when the response list is longer than the premise list, when responses can be used more than once, and/or when some of the responses do not match any of the premises.

As summarized in Table 1, there are a number of advantages to matching formats. The compact format of matching allows for large amounts of factual information to be included in the test relative to the amount of test space available. Matching provides a quick means for testing associations, such as terms, events, places, dates, causes, results, and symbols. It is a good format for testing who, what, when, and where types of information. Matching items are

---

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, Ph.D., is in the School of Education at the University of Wyoming.

This article is the fourth in a series on teacher-designed examinations contributed by Gary and Kay Holt.

Journal of Pharmacy Teaching, Vol. 2(2) 1991

TABLE 1. Advantages of Matching Questions

1. Can measure large amounts of information

   Because of the compact format of matching items, a considerable amount of information can be covered in a relatively small amount of space.

2. Provide a means for quickly testing associations such as words/definitions, events/places, results/causes, concepts/symbols, authors/works

3. Good format for testing who, what, when, and where types of information

4. Items relatively easy to construct and score

5. Can use responses more than once

6. Reduce guessing

   The more options (i.e., responses) that are provided, the less chance there will be for students to guess correctly.

7. Useful when the teacher wants to test for student mastery of a large number of facts, ideas, principles, or definitions

relatively easy to construct and score, and they help reduce the tendency to guess.

Table 2 summarizes the disadvantages associated with matching. For example, even though matching items are relatively easy to construct, it can still be difficult to write good items. Some subject matter simply does not lend itself well to matching formats. Also, matching items may not be well suited to the measurement of higher cognitive skills. Instead, most matching items measure knowledge (i.e., memorization) level. Finally, most commercial answer sheets cannot accommodate more than five response options. This can pose a problem if the instructor prefers to use an electronic test scoring device.

Table 3 summarizes guidelines for constructing matching items. The premise and response lists should be as homogeneous as possible. Each list should be related to one subject. Mixing topics makes it easier for the student to categorically eliminate certain choices. Thus, even with limited knowledge, the student may perform well on the test merely by a process of elimination.

It is important to provide directions that are complete and clear. Instructors should avoid the use of specific determiners which clue the student as to the type of answer required. It is important to design the premise and response lists in such a way as to reduce random searching. This is best accomplished by placing the lists in some type of logical order, for example, alphabetical order. Then the students should be informed that this order has been used.

If possible, it is best to place the shortest items in the response column. This is a time-saving technique. The students can read the longer premise, then search more quickly through the shorter response items. The length of the premise and response lists should be limited to 10 items in the premise list and 15 items in the response list. And it is always a good idea to have more items in the response list than in the premise list to reduce the opportunity for guessing by elimination.

Both lists should always appear in their entirety on the same page. Too much time is wasted when students have to flip from one page to another to search for response options.

TABLE 2. Disadvantages of Matching Questions

1. Good items can be difficult to construct.

   Sometimes the subject matter does not lend itself well to matching formats, or homogeneous premises and responses may be hard to find.

2. Matching items are usually not well suited to measuring higher cognitive skills.

   Most items tend to measure the knowledge (i.e., memorization) level. Items often ask students to address information that is relatively trivial in nature, although it is certainly possible to construct matching items that measure more complex information.

3. Most commercial answer sheets for use with electronic scoring devices can only accommodate five options. Thus, it is often difficult to score matching sections electronically.

ESSAY EXAMS

An essay question is one for which the student supplies a composed response, usually in one or more sentences, rather than selecting a response from a list of alternatives. Essay exams differ from short-answer tests in degree rather than in kind. They usually allow greater freedom of response to questions and require more writing. Aside from the oral exam, the essay is the oldest test format still in use today, having been stressed by Horace Mann in about 1845. It is a rewarding and useful testing format for all concerned. If essay exams are used to help students learn more by thinking about classroom materials in new ways, their pedagogical benefits become unquestionable.

There are two types of essay exam. The first is called a restricted response, which requires the student to supply a brief answer to a limited question. The second is called an extended response and requires the student to write an extensive answer in which information deemed by the student to be relevant to the problem must be selected, organized, and integrated. Essay exams are more appropriate when course objectives call for the demonstration of higher cognitive skills rather than simple recognition. They also seem to be more applicable in smaller classes and when the test will not be reused.

Table 4 summarizes the advantages of essay exams. They allow students to have freedom to respond within specified limits. They are generally easier to prepare, since only a few questions are usually needed for each exam. The opportunity for guessing is reduced because students do not select the answer from a list of alternatives that have been provided. Essay questions can be written to cover most subjects, and they promote a more desirable method of study and preparation for the exam.

Table 5 summarizes the disadvantages that have been proposed for essay test items. The major problem is the difficulty of grading them. Because students have greater freedom to write, grading can become somewhat subjective. Studies have indicated that there can be a considerable lack of consistency between educators in the same discipline. For example, experienced teachers tend to grade harder. Two teachers may grade the same paper as much as two letter

TABLE 3. Construction Guidelines for Matching Items

1. Make the lists of premises and responses as homogeneous as possible.

   Each list should be confined to one type of subject. Mixing a variety of topics makes it easier for the student to categorically eliminate certain choices. Thus, even with limited knowledge they may answer questions correctly. Generally speaking, every response should be a viable option for every premise.

2. Provide complete directions.

   Directions should tell the student how to respond. This should include where responses should be written, whether responses can be used more than once, a description of what is contained in the columns, and the basis upon which matching should be made.

3. Avoid specific determiners.

   Avoid clues in the premise column that identify certain responses (e.g., gender terms, plural endings).

4. Design the premise and response lists so that random searching is reduced.

   This is best accomplished by arranging both lists in a logical order (e.g., alphabetical or numerical order--depending upon what is contained in the lists). Students should be informed that this has been done to prevent them from seeking the system that has been used.

5. Whenever possible, place the shorter items in the response column.

   This reduces the time required to search the response column.

6. Limit the number of items within each set.

   When the lists are too long, the students will spend inordinate amounts of time reading and searching, even when they know the answer. It is usually wise to limit the premise list to 10 items and the response list to 15 items.

7. Include more responses than premises.

   This is especially true if each response can be used only once. A process of elimination can occur if items are matched one-to-one.

8. Place both lists, in their entirety, on the same page.

   Time is wasted if students must flip from one page to another to search for the responses.

TABLE 4. Advantages of Essay Test Items

1. Allow students to express their ideas with relatively few constraints

2. Practical for testing smaller numbers of students

3. Can measure divergent thinking and higher cognitive skills

   These items allow for unconventional, creative, and relatively rare responses. A good essay item should require the student to select relevant information, organize it, and express it clearly. These items can require the student to apply, analyze, synthesize, or evaluate information. Not only must students know the facts, but they must be able to present them in logical ways to support their arguments. Essay items can force students to think about course materials in new ways.

4. Generally easier to prepare, since only a few questions are usually needed for each exam

5. Reduce opportunity for guessing

   Essay items do not involve selection of a response from a list of options. Instead, the student must supply the response by recall.

6. Can be written in almost all subject fields

7. Generally require a more desirable method of study

   Students tend to learn ideas, concepts, and the general flow of events rather than a mass of isolated facts. Better than any other test format, essay exams reward the ability to think, organize, and apply knowledge. It is thought that this can lead to better understanding and greater retention.

grades apart, and the same teacher may grade a paper differently at one time than another. Furthermore, writing styles, spelling, the student's previous performance in school, and the degree to which the teacher likes the student can affect the final grade. Other disadvantages exist. Essay exams tend to cover less material. Students may successfully bluff their way to a higher grade if they happen to be accomplished writers. Essay questions only demand higher cognitive skills when the questions have been appropriately prepared. Finally, essay test items are very writing oriented. This can be a particular disadvantage if the majority of the students' time is devoted to the mechanics of writing rather than to the thought behind the response. On objective test items, very little time is spent in writing responses, so the students spend more time thinking about responses.

Guidelines for the writing of essay tests are outlined in Table 6. These are important because many of the problems associated with essay tests involve inadequate teacher preparation. Specifically, teachers often neglect to specify exactly what they want their students to do. It is important to specify limitations. Students should be told how long responses should be and the point value for each question. This information helps the student plan the response that will be provided. The question should indicate the extent and depth of the desired response, and if space limitations are desired (e.g., a certain number of pages), this should be included as well. Similarly, the freedom of the response should be indicated. This can be a problem because both shorter and longer answers can be valid. Even so, the test item should indicate the nature of the response that is expected by the teacher.

It is better to design essay test items that are clear, specific, and to the point and that require shorter, specific answers than to write test questions that ask the students to relate everything that is known about a particular topic. The latter approach can result in responses that are vague and difficult to score. Furthermore, the question should emphasize higher cognitive skills in the response rather than superficial or trivial information.

Students should be required to answer the same questions. Educators often write several questions, then have the students select only two or three to answer. This approach is valid only if the ob-

TABLE 5. Disadvantages of Essay Tests

1. Difficult to score objectively

   Students have greater freedom to express themselves, and this creates problems regarding objective evaluation of responses. Additionally, long, complex essay questions are more difficult to score than shorter, more restricted ones.

   Studies indicate that teachers with more experience and expertise tend to grade harder. And it is not uncommon for different teachers to score the same essay examination with a difference of as much as two letter grades due to the different criteria teachers use for grading.

   Halo effects may also pose a problem. These involve positive and negative influences of irrelevant factors on the scoring of the exam (e.g., writing styles, spelling, performance on earlier questions, personality of the student).

2. May measure only limited aspects of student knowledge

   Since each question takes a relatively long period of time for a response, fewer questions can be included on the exam.

3. Time-consuming for both teacher and students

   Students may spend considerable amounts of time answering only a few questions. Teachers may devote many hours to reading lengthy responses and scoring responses.

Studies have shown that objectivity is increased more by increasing the number of items than by allowing greater freedom in responding to fewer items.

4. Subject to bluffing

   Although guessing is reduced, bluffing can occur. Poorly prepared students or those with a "gift for gab" may write something (often lengthy), even if the response is poorly related to the question.

5. Often require little more than rote memorization

   Essay questions demand a demonstration of higher cognitive skills only when they have been designed to ask for this type of response. In practice, many essay questions require no originality, and many emphasize the lengthy enumeration of facts or trends.

6. Place an emphasis on writing

   Much of the time allotted to answering an essay question is devoted to the mechanics of writing, with relatively little time to think about content. On more objective types of test items, little time is spent in writing; therefore, more time can be devoted to thinking about responses.

TABLE 6. Guidelines for Writing Essay Items

1. Specify limitations.

   Specify space, wording, or time limitations. The extent and depth should also be indicated if these parameters will affect grading.

2. Determine the amount of freedom in responses.

   Frame each question in such a way that the students understand the nature of the response that the teacher expects.

3. Make each essay item relatively short and increase the number of items.

   Long essay questions that require students to relate all that they know about some topic will usually result in vague responses that are difficult to score. By designing a larger number of more specific items, it is possible to sample student knowledge more comprehensively and systematically. This latter approach also increases ease of scoring.

4. Tell the students how many points each question is worth.

   This allows the students to appreciate the priorities being placed upon the various questions and directs the them to use test time accordingly.

5. Write essay items that emphasize higher cognitive skills.

   These include questions that emphasize applications, analysis, synthesis, and evaluation.

6. Write essay questions that are clear, specific, and to the point.

   A common problem with essay questions is that they are too vague and general. This encourages student responses that are vague and general.

7. Have all students answer the same questions.

   Teachers often write several questions and give students an option as to which they will answer. However, this approach should not be used if the objective of the test is the demonstration of knowledge. In essence, allowing students to choose the questions they want to answer means that they are not taking the same exam. Because the questions will likely differ in complexity, poorly prepared students could actually opt for easier questions and perform better on the exam than better prepared students who opt for more difficult questions. When subject matter is important, all students should answer the same questions.

8. Try to design tests that adequately sample the subject matter being tested.

   Because essay tests typically cover less of the course content, it is possible that a well-prepared student might be given a question in his/her one area of deficiency. Thus, essay tests can underestimate actual student achievement. For this reason, it is important to design questions that reflect important aspects of the course content.

9. Establish appropriate time limits.

   There is a tendency to allow too little time for taking essay tests. This occurs because it is more difficult to estimate the time required for essay tests. Forcing students to rush can result in a deterioration in the legibility of handwriting and organization of responses. Ultimately, this presents a problem in evaluation and scoring.

TABLE 6 (continued)

10. Questions should be related to the course objectives.

11. Design items that are appropriate for the students' background.

12. Once the test is written, set it aside for a few days or weeks.

    Rereading it later allows the teacher to more objectively evaluate it before giving it to students.

jective of the test is to demonstrate writing skills rather than acquisition of information. Thus, students' selections may not be representative of their actual knowledge base. If the objective of the exam is to measure comprehension or understanding of course content, this approach should not be used. In essence, allowing students to choose the questions they want to answer means that they are not taking the same test. Because the questions will likely differ in complexity, poorly prepared students could actually opt for easier questions and perform better on an exam than better-prepared students who opt for more difficult questions. When the subject matter is important, all students should answer the same questions.

Every effort should be made to present an adequate sample of the subject matter being tested. Because the typical essay test has fewer questions, the test may represent only a small portion of the course content. It is possible that generally well-prepared students might be asked the few questions for which they are not well prepared. In such a case, the test scores would not reflect the students' actual preparation.

It is important to give some thought to the time limits of the exam relative to the number of questions that will be asked. It is more difficult to determine how much time students will need for essay exams than it is for objective test questions. Additionally, it is difficult to predict the speed at which students will write, the time required to select and organize information, and the impact that fatigue can have. The tendency is to allow too little time for taking essay exams. This can cause students to rush, possibly resulting in a deterioration in the legibility of handwriting and organization of responses. Subsequently, this presents a problem in evaluation and scoring. When in doubt, allow too much time rather than too little, since the content validity of the test is improved when speed is not a factor.

Finally, most educators set the test aside for a few days once it has been written. Rereading an exam at a later date allows the teacher to evaluate it more objectively before giving it to students.

Because the grading/scoring is the greatest problem with essay exams, this topic warrants additional consideration. Table 7 summarizes guidelines suggested by education scholars. In general, grading/scoring is a concern to educators because of the time re-

TABLE 7. Methods for Scoring Essay Items

1. Analytical Method

   a. Design a model answer for each question.
   b. Analyze it and identify its component parts.
   c. Assign point values to the component parts.
   d. Compare student responses to the model answer and score accordingly.

2. Rating Method (Global Method or Sorting Method)

   a. Design a model answer for each question.
   b. Compare student responses to the model answer.
   c. On the basis of the wholeness of each student response, categorize student responses (e.g., good, average, poor). The number of category divisions normally ranges from 1 to 9.
   d. Reread student responses a second time and, if needed, a third. Compare these reevaluations with the first evaluation.
   e. When each response seems to have been appropriately categorized, score accordingly.

quired, the subjectivity involved in response evaluation, halo effects, and other problems. Because of the subjectivity involved, teachers must be able to defend the grades they have given. This requires the development of specific guidelines for why a certain grade was given and exactly what could have been included in the response to achieve a higher grade. Fortunately, studies indicate that the grading of essay exams can be fairly reliable if teachers follow certain procedures.

There are two basic methods for scoring essay items. The first is called the analytical method, or point method. Basically, it involves developing a perfect response, identifying the components of that response, assigning point values for each component, then comparing the student's response to this perfect response. Students are assigned points relative to the number of component parts of the perfect response that they included in their responses.

The second method is called the rating method, global method, or sorting method. Again, it involves the development of a perfect or ideal response. However, in this case there are no component parts of the response. Rather, the wholeness of the response is considered. Student responses are read and compared to this global, ideal response. On the basis of the wholeness of the students' responses, the teacher places the student responses into certain categories (e.g., good, average, poor). The number of divisions can vary, but the final score is based upon the category into which the student has been placed.

Table 8 summarizes additional suggestions for reducing subjectivity in grading. For example, regardless of the method of grading selected, a model answer should be prepared as a guide for grading. This may be prepared by constructing a model answer from the course materials or by reading a small number of the student responses, without scoring them, to obtain an idea of how the students performed.

It is best to somehow conceal the students' names while the papers are being graded and also to conceal the scores of previously graded test items. This helps to reduce the possibility of negative or positive halo effects that occur during grading. It is also recommended that teachers read and evaluate each student's answer to the

TABLE 8. Additional Guidelines for Grading Essay Tests

1. Evaluate anonymous items.

   If possible, conceal the identity of the student. Halo effects can be reduced when items are graded based upon what is written without considering who prepared the response.

2. Read and evaluate each student's answer to the same question before moving to the next question.

   This approach results in more consistent grading, minimizes halo effects, and has been found to speed the grading process.

3. Keep scores of previously scored items out of sight.

   The student's performance on previous items can influence grading on subsequent items. Concealing previous scores helps to reduce this negative halo effect.

4. Put comments on the paper to let students know the basis upon which evaluations were made.

5. Use multiple grading.

   This can be accomplished by having the same teacher grade the papers on different occasions or by having more than one teacher evaluate the papers. Scores from each grading are compared. If they are the same or sufficiently similar, this is the score that is assigned. If a sufficient discrepancy exists, the paper should be reevaluated.

6. Prepare a model answer for each question.

   The model should include the points a student would include, as well as point values for each question or subpart of each question.

7. Read a small number of papers without scoring them to obtain a general idea of student performances.

   This allows for a realistic set of grading standards in the event that teacher expectations are different from actual student responses.

8. If spelling, penmanship, grammar, and writing style are to be graded, they should be considered independent of the content.

   How students write should be considered apart from what they write.

9. Decide on a policy for handling irrelevant ("empty") answers or incorrect responses.

   Students often use many words to say very little. They may hope that the teacher will find answers that are not actually there, or that the teacher will stumble onto a few key words that give the impression of intellectual accomplishment. Students should be told how irrelevant, incorrect, or illegible responses will be handled, and the teacher should adhere to the principles that are established.

same question before moving to the next question. This results in more consistent grading throughout the class.

Teachers should make comments on the papers so that students know how the grades have been determined. Multiple grading can help to ensure objectivity. This can be accomplished by regrading each paper at a later time or by having another instructor grade the papers. Similarly, if spelling, penmanship, grammar, and writing style are to be considered, they should be graded independent of the content.

Finally, a policy should be designed up front for handling irrelevant ("empty") answers or incorrect responses. A common problem with essay exams is that students may say little in a great many words. They may hope that the teacher will find answers that are not actually there, that teachers will not try to struggle through sloppy penmanship, or that teachers will stumble onto a few key words that give an impression of intellectual accomplishment that does not actually exist. When the teacher emphasizes that irrelevant, incorrect, or illegible responses will be handled in a particular manner, the students are discouraged from such tactics.

OPEN-BOOK TESTS

Open-book tests allow students to use their books or other materials during the examination. These are commonly used when teachers want to emphasize higher cognitive skills (e.g., the application of formulas or facts) rather than merely have students demonstrate their memorization skills. If the purpose of the test is one of measuring knowledge, the use of the traditional closed-book test is probably more appropriate. In such cases, the use of open-book tests may demonstrate nothing more than the student's ability to use the textbook and other reference materials. However, if the purpose of the test is to demonstrate the student's ability to apply knowledge, the traditional closed-book examination may not be the most desirable method of testing. For example, open-book exams can be quite useful in allowing students to apply their knowledge to daily practice or living scenarios. In such cases, the examination should simulate the natural situation as much as possible, allowing the student access to tools and aids that would normally be available.

Open-book tests tend to be more dependent upon innate understanding, self-expression skills, and reasoning and evaluative abilities.

The advantages and disadvantages of open-book tests are summarized in Tables 9 and 10, respectively. Table 11 provides tips for constructing open-book tests, and Table 12 provides guidelines for the administration of this type of exam.

TAKE-HOME EXAMS

The take-home exam is an extension of the open-book test. Students are allowed to work at their leisure and at their most comfortable rate of speed, using whatever resources are available. It is a useful approach when students need additional resources not normally available in the classroom. Homework assignments are similar projects, but they are not usually graded as exams, although they can be. The advantages and disadvantages of take-home exams are summarized in Tables 13 and 14, respectively.

TABLE 9. Advantages of Open-Book Tests

1. An excellent type of exam when appropriately designed and used

2. Possibly reduce cheating and test anxiety

3. Promote learning

   They encourage learning by allowing students to apply their own thoughts and ideas in conjunction with reference materials. They also promote correct use of reference materials.

4. Realism

   Open-book exams are thought to more closely parallel reality in that many work situations require the use of reference materials. These exams require that students gather needed information; evaluate and synthesize it; and then express it in a logical, coherent manner.

TABLE 10. Disadvantages of Open-Book Tests

1. More difficult to grade and require the teacher to have an excellent understanding of the material to evaluate student responses adequately

2. Require increased grading time (as compared to objective test items)

3. Difficult to discriminate better students from average students when trying to assign an objective grade to a subjective performance

4. Grades sometimes higher

   This is not a true disadvantage, since the issue should be what students accomplish and not what scores they earn. However, many educators view higher grades as a disadvantage.

TABLE 11. Guidelines for Constructing Open-Book Tests

1. Avoid textbook terminology.

   Copying questions or problems verbatim from a textbook does not require students to think about the problem involved.

2. Design the problems or test items so that reference materials will be used in the same manner as they will in the students' natural work environment.

3. Discourage the overuse of resources.

   Having too few questions allows students to spend inordinate amounts of time looking in the reference materials. Having an appropriate number of questions encourages more initial preparation on the part of students, so they are not totally dependent upon the references. However, it is important to warn the students prior to the exam that adequate preparation is required due to the number of questions on the exam.

4. Design questions whose answers cannot be easily looked up or copied.

   Test questions should require the application of reference materials and the demonstration of more complex thought processes.

TABLE 12. Administration of Open-Book Tests

1. If open-book tests are to be used, students should be forewarned.

   The type of test that students will be given affects the manner in which they prepare.

2. Inform the students prior to the test which reference materials will be available to them.

TABLE 13. Advantages of Take-Home Exams

1. Can be valuable for students when used as an exercise or teaching device to indicate student strengths, weaknesses, and methods of improvement (rather than for grading)

2. Do not take up class time

TABLE 14. Disadvantages of Take-Home Exams

1. Most difficult essay-type responses to grade

   Take-home test responses tend to be long, complex, and the most varied of all essay-type exams.

2. Possible that students will not do their own work

3. No opportunity for students to ask legitimate questions about test items or procedures involved