

IV. Student Assessment

Chapter 12 - Testing and Assessment Issues

Assessment of student achievement is an important part of the teaching and learning process. Given at the beginning of a course, assessments help you know where to begin and/or identify areas of remediation that must be addressed. Frequent assessments during the course help you and your students see the progress of learning and help identify problematic areas where students need more help or time. Given at the completion of instruction, assessments tell you how much has been learned by the end of a unit, by mid-semester, or by the end of the term. They provide the basis for making judgments on the grades to assign each student. This chapter provides an overview of assessments, with a focus on tests or examinations typically given in paper/pencil format or on the computer.

- **Types of Learning Assessments**
- **Paper/Pencil or Computer Examinations (Tests)**
- **Step-by-Step Guidelines for Creating Tests**
- **Constructing Performance Tests**
- **General Tips about Testing**
- **Helping Students Learn from Tests**
- **Using Item Analysis to Test Your Test**
- **Cheating on Tests**
- **Alternative Methods of Assessment**
- **Resources on Testing**

Types of Learning Assessments

- **Examinations**
 - Open-ended such as essay and short-answer.
 - Limited-choice such as multiple choice, sentence completion, fill-in-the-blank, matching, true-false.
 - Usually provided in pencil/paper format, sometimes involving scan response sheets or administered on a computer.
- **Written or Constructed Student Creations**
 - Reports, papers, projects, products.
 - Usually done outside of class and involving research or reviews of a variety of information sources. Final products are assembled for submission.
- **Performances**
 - Demonstrations, events, presentations.
 - Students demonstrate skills and knowledge in simulated or authentic conditions. May focus on psychomotor skills but can also heavily involve cognitive skills and judgments, such as in counseling performances.

Paper/Pencil or Computer Examinations (Tests)

Purposes of Effective Assessments

- **Intended learning outcomes:** Measure what students should know and/or be able to do to show they have mastered the learning outcomes.
- **Knowledge and skills included in the instruction:** Measure the amount and quality of student ability to use information, examples, practices, and other related activities provided during instruction.
- **Enable generalization:** Allow inference from skills and knowledge tested that students have mastered the full range of skills and knowledge and the essential or key content points.

Alignment of Course Components

Assessments should align with the course objectives or specified learning outcomes and the content and activities included in the class. The appropriate range of content should be assessed, as well as the key points covered during instructional activities. The assessment should allow the student to demonstrate his or her knowledge and skill in the subject area.

Constructing Tests

Limited-Choice vs. Open-Ended Questions

The term “limited-choice” is used here to describe test questions that require students to choose one or more given alternatives (multiple choice, true/false, matching), and “open-ended” is used to refer to questions that require students to formulate their own answers (sentence completion, short answer, essay).

Deciding Which Type of Test Questions to Use

Whether it is better to use open-ended or limited-choice test items depends on the circumstances and on the goals of the test. Each type of test has its own sets of strengths and weaknesses. The advantages and disadvantages of the two main categories of test items are discussed below in terms of the various issues that are often considered when a test is being developed.

Table 1 - Comparison of Limited-Choice and Open-Ended Tests

Issue	Limited Choice	Open-Ended
Level of learning objective	Recall, comprehension	Problem solving, synthesizing
Content coverage	Wider sample	Greater depth
Practice and reward of writing and reading skills	No	Yes
Reward of creativity and divergent thinking	No	Yes
Feedback to instructor and student	Limited but fast	Thorough but slow
Length of exam (time to complete)	Short	Long
Size of class	Larger	Smaller
Reliability in grading	Very reliable	Requires work to become reliable
Exam construction and grading time	Long/short	Short/long
Test reusability	High	Low
Prevention of cheating	Low	High

Level of Learning Objective

In principle, both limited-choice and open-ended items can be used to test a wide range of learning objectives. In practice, most people find it easier to construct limited-choice items to test recall and comprehension, while open-ended items are used to test higher-level learning objectives, but other possibilities exist. Limited-choice items that require students to classify statements as fact or opinion go beyond rote learning, and focused essay-questions can easily stay at the recall level.



Related Chapter

Chapter 2

Determining Learning Objectives, for discussions of the different levels of learning outcomes.

Content Coverage

Since more limited-choice than open-ended items can be used in exams of the same length, it is possible to sample more broadly over a body of subject matter with limited-choice items.

Scoring and Grading

Limited-choice exams allow faster and more consistent scoring than open-ended exams. Open-ended exams require individual review and judgment of student responses and, therefore, take longer to score and may be scored more subjectively, even by the same reviewer. Unless graders are available, it is very difficult to give long open-ended exams and provide timely feedback in a high-enrollment course. Exams that consist mainly of limited-choice items are usually more practical under these circumstances.

Test Construction

Constructing good limited-choice exams takes much longer than open-ended exams because items must contain the pertinent information students need to answer the question, a set of appropriate distracters, and the correct answer. In addition, none of the information should include clues that point to the correct answer or be written in a way that confuses student reading and interpretation of the item. Open-ended exams, on the other hand, may be constructed quickly and easily because they usually consist of one or two direct statements or questions asking students to respond in writing. The chances of providing unwanted clues are greatly reduced, and there is no opportunity to confuse students with distracters that may be too close to the right answer.

Length of Exam and Student Response Time

Whether using limited-choice or open-ended exams, instructors should consider how much time students might need to respond thoughtfully to the test items. One frequent complaint from students is that they knew the material but there were so many items that they could not answer every question or could not take time to provide thoughtful answers. Asking a colleague or a graduate student who has taken the course to complete the test may give you some estimate of how many items can be completed during the time provided.

Reusability of Exam Items

In general, exams consisting of a large number of limited-choice items are easier to reuse than those consisting of only a few essay questions. More items are more difficult for students to remember and transmit to those who will take the exam later (if the printed exam does not get into circulation). If a large item bank is built and different exams can be randomly generated from the same pool of questions, limited-choice items are highly reusable.

Prevention of Cheating

Limited-choice exams provide more opportunities for cheating than do open-ended exams since single letters or numbers are far easier to see or hear than extensive text. Cheating on limited-choice items can be minimized in several ways, such as using alternative test forms and controlling students' seating arrangements.

Writing Limited-Choice Test Questions

In the discussion of limited-choice items below, the term “stem” is used to refer to the part of the item that asks the question. The terms “responses,” “choices,” “options,” and “alternatives” are used to refer to the parts of the item that will be used to answer the question.

Example

Stem: Who is the author of *Jane Eyre*?

Responses:

- A) Emily Bronte
- B) Charlotte Bronte
- C) Thomas Hardy
- D) George Elliot

Multiple Choice Items

Advantages - Multiple-choice items are considered to be among the most versatile of all item types. They can be used to test students' ability to recall facts as well as their understanding and ability to apply learning. Multiple-choice items can also provide an excellent basis for post-test discussion, especially if the discussion addresses why the incorrect responses were wrong as well as why the correct responses were right.

Disadvantages - Unfortunately, good multiple-choice items are difficult and time-consuming to construct. They may also appear too discriminating (picky) to students, especially when the alternatives are well constructed, and open to misinterpretation by students who read more into questions than is there.

Suggestions for Constructing Multiple-Choice Items

Concerns about the general construction of questions

- Use negatively stated items sparingly. When they are used, it helps to underline or otherwise visually emphasize the negative word. Never use the word “not” in a multiple-choice question.
- Be certain there is only one best or correct response to the stem.
- Keep the number of alternatives at five or fewer. Beyond five alternatives, poor alternatives are likely.
- Randomly distribute correct responses among the alternative positions so that there are no discernible patterns to the answer sequence (e.g., ABBABBABB). Try to have a nearly equal proportion of As, Bs, Cs, etc., as the correct answers.

Concerns about the construction of the stem portion of the question

- Use the stem to present the problem or question as clearly as possible.

- Use direct questions rather than incomplete statements for the stem.
- Include as much of the item as possible in the stem so that alternatives can be kept brief. However, when applying definitions, it is recommended you place the terms in the stem and use the definitions as options, although this makes the questions rather long.

Concerns about the construction of the responses or options of the question

- List options on separate lines rather than including them as part of the stem, so that all options can be clearly distinguished.
- Keep all alternatives in a similar format (i.e., all phrases, all sentences, etc.).
- Be certain that all options are plausible responses to the stem. Poor alternatives should not be included just for the sake of having more options.
- Check all choices for grammatical consistency with the stem.
- Try to make alternatives for an item approximately the same length. Making the correct response consistently longer is a common error.
- Use misconceptions students have displayed in class, or errors commonly made by students in the class, as the basis for incorrect alternatives.
- Use “all of the above” and “none of the above” sparingly since students, on the basis of incomplete knowledge, choose these alternatives often.

- Use capital letters (A, B, C, D, E) as response signs rather than lower-case letters (“a” gets confused with “d” and “c” with “e” if the quality of the typeface or duplication is poor).

True/False Items

Advantages - True/false items are relatively easy to prepare since each item comes rather directly from the content. They offer the instructor the opportunity to write questions that cover more content than most other item types since students can respond to many questions in the time allowed. They are easy to score accurately and quickly.

Disadvantages - True/false items, however, may not give a true estimate of the students’ knowledge since students have a 50/50 chance of guessing the correct answer. They are very poor for diagnosing students’ strengths and weaknesses and are generally considered to be “tricky” by students. Since true/false questions tend to be either extremely easy or extremely difficult, they do not discriminate between students of varying ability as well as other types of questions.

Suggestions for Constructing True/False Items

- Keep language as simple and clear as possible.
- Use a relatively large number of items (75 or more when the entire test is T/F).
- Avoid taking statements verbatim from the text.
- Be aware that extremely long or complicated statements will test reading skill rather than content knowledge.

- Require students to circle or underline a typed “T” or “F” rather than to fill in a “T” or “F” next to the statement. This allows scorers to avoid having to interpret confusing handwriting.
- Avoid the use of negatives, especially double negatives. Never use “not.”
- Avoid ambiguous or tricky items.
- Be certain that the statements used are entirely true or entirely false. Statements that are either partially true or partially false cause unnecessary ambiguity.
- Use certain key words sparingly since they tip students off to the correct answers. The words “all,” “always,” “never,” “every,” “none,” and “only” usually indicate a false statement, whereas the words “generally,” “sometimes,” “usually,” “maybe,” and “often” are frequently used in true statements.
- Use precise terms, such as “50% of the time,” rather than less precise terms, such as “several,” “seldom,” and “frequently.”
- Use more false than true items, but do not exceed their use more than 15%. False items tend to discriminate more than true items.
- Avoid patterns in answers such as “all true,” “all false,” or “alternation.”

Matching Items

Advantages - Matching items are generally quite brief and are especially suitable for who, what, when, and where questions. They can, however, be used to have students discriminate among, and to apply concepts. They permit efficient use of space when there is a number of similar types of information to be tested. They are easy to score accurately and quickly.

Disadvantages - Among the drawbacks of matching items are that they are difficult to use to measure learning beyond recognition of basic factual knowledge, and they are usually poor for diagnosing student strengths and weaknesses. Matching items are appropriate in only a limited number of situations, and they are difficult to construct, since parallel information is required.



Example

Notice the relative width of the columns in the “Cities of the World Quiz” on the following page. Also notice that the directions tell the learner what to do and answer possible questions about the format of the quiz.

Cities of the World Quiz

Directions: A description of, or fact about, a major city in the world appears as part of the numbered question. The city names are listed on the right. Write the capital letter corresponding to the correct city in the list on the line corresponding to each question. You may use cities from the list more than once. Some cities may not be described at all.

- | | |
|--|---|
| ___1. The Seine river divides this city into two famous banks. | A. Kyoto |
| ___2. This obscure Roman fortress city suffered four major fires on its way to becoming capital of an empire larger than Rome. | B. Madison
C. London |
| ___3. The capital city of the Island of Taiwan | D. Paris |
| ___4. Once a capital of the Roman empire, this city became the capital of the Eastern Orthodox faith. | E. Tallahassee |
| ___5. The tallest building in the world is located in this city. | F. Chicago |
| ___6. Called the “City of Big Shoulders,” this city was once home to the world’s largest stockyards. | G. Rome |
| ___7. Home city to the Statue of Liberty | H. Lisbon |
| ___8. Located on a continental divide, this city’s builders reversed the direction of flow of water in the City’s river. | I. Moscow
J. Taipei |
| ___9. This city was once the winter capital of Japan. | K. Constantinople |
| ___10. The Kremlin is located in this city. | L. Beijing
M. New York
N. Dubai
O. Cape Town |

Suggestions for Constructing Matching Items

- Use only homogeneous material in a set of matching items, i.e., dates and places should not be in the same set.
- Use the more involved expressions in the stem, and keep the responses short and simple.
- Supply directions that clearly state the basis for the matching, indicating whether a response can be used more than once and stating where the answer should be placed.
- Be certain there are never multiple correct choices for one premise, although a choice may be used as the correct answer for more than one premise.

- Avoid giving inadvertent grammatical clues to the correct choice by checking that choices match each other in terms of tense, number, and part of speech.
- Arrange items in the response column in some logical order - alphabetical, numerical, chronological - so students can find them easily.
- Avoid breaking a set of items (premises and choices) over two pages.
- Use no more than 15 items in one set.
- Provide more choices than premises to make “process-of-elimination” guessing less effective.
- Number each premise for ease in later discussions.
- Use capital letters for the response selections rather than lower case letters. Insist that a capital letter be written in the area where the answer is placed.

Writing Open-Ended Test Questions

Completion Items

Advantages - Completion items are especially useful in assessing mastery of factual information when a specific word or phrase is important to know. They preclude the kind of guessing that is possible on limited-choice items, since they require a definite response rather than simple recognition of the correct answer. Because only a short answer is required, completion items require a wide sampling of content.

Disadvantages - Completion items, however, tend to test only rote and repetitive responses, and they may encourage a fragmented study style since memorization of bits and pieces of information can result in higher test scores. They are more difficult to score than forced-choice items, and scoring often must be done by the test writer since more than one answer may have to be considered correct.

Suggestions for Constructing Completion Items

- Use original questions rather than taking questions from the text.
- Provide clear and concise cues about the expected response.
- Use vocabulary and phrasing that comes from the text or class presentation.
- Provide explicit directions, when possible, as to what amount of variation will be accepted in the answers.
- Avoid using a long quote with multiple blanks to complete.
- Require only one word or phrase in each blank.
- Facilitate scoring by having the students write their responses online arranged in a column to the left of the items.
- Ask students to fill in only important terms or expressions.

- Avoid providing grammatical clues to the correct answer by using “a,” “an,” etc., instead of specific modifiers.
- Assign much more credit for completion items than for T/F or matching items.

Essay/Short-Answer Items

Advantages - Short-answer items, those limited to fewer than five full sentences, are interchangeable with completion items. Essay items, on the other hand, allow expression of both breadth and depth of learning, and encourage originality, creativity, and divergent thinking. Written items offer students the opportunity to use their own judgment, writing styles, and vocabularies. They are less time-consuming to prepare than any other item type.

Disadvantages - Unfortunately, tests consisting only of written items permit only a limited sampling of content learning due to the time required for students to respond. Essay items are not efficient for assessing knowledge of basic facts, and they provide students more opportunity for bluffing and rambling than do limited-choice items. They favor students’ good writing skills and neatness, and they are pitfalls for students who tend to go off on tangents or misunderstand the main point of the question. The main disadvantage is that essay items are difficult and time-consuming to score and are subject to biased and unreliable scoring.

Suggestions for Constructing Essay/Short-Answer Items

- Use novel problems or material whenever possible, but only if they relate to class learning.
- Make essay questions comprehensive rather than focused on small units of content.
- Require students to demonstrate command of background information by asking them to provide supporting evidence for claims and assertions.
- Provide clear directions about expectations.
- Allow students an appropriate amount of time. It is helpful to give students some guidelines on how much time to use on each question, as well as the desired length and format of the response, e.g., full sentences, phrases only, outline, and so on.
- Inform students, in advance, about the proportional value of each item in comparison to the total grade.
- Keep grading in mind while creating the questions. Jot down notes of what you expect to see in student answers that help identify mastery of the subject matter.

Step-by-Step Guidelines for Creating Tests

- Determine which types of items are best for the testing situation, and then write them.
- Write explicit directions for the test sections indicating credit on each section.
- Organize the layout (group like items together; start with easy items; number the items).
- Make the answer key.
- Review patterns of responses (avoid sequences such as ABABCABABC).
- Use alphabetic, chronological, or numerical sequences to determine how response choices are organized, all of which will help you avoid getting into a pattern of responses.
- Develop a scoring method or scheme.
- Weight test points according to types of item, learning assessed, and student effort involved.
- Score test papers anonymously.
- Observe student confidentiality.
- **Review the final product.**
 - Are the items concise?
 - Have inadvertent clues been avoided?
 - Do the number of items written for each objective, or topic area, represent the emphasis placed on them during instruction?

- Do the difficulty levels of the items seem appropriate?
- Is the length of the test appropriate?
- Are the test items readable (understandable)?
- Have spelling errors and typos been corrected?
- Ask an “outside reviewer” available to critique the test for content, difficulty level, and timing.
- Make final changes and then duplicate the test.

Constructing Performance Tests

Advantages - The truest measure of whether a learner is able to do something is to watch the learner do it. Performance tests provide an opportunity to make this kind of measurement. As long as you are able to observe the student’s performance of a prescribed task, confidence in the student’s ability is validated.

Disadvantages - While you may see a performance successfully completed, normalcy of the performance situations is not assured. Also, know that failure to perform may not show what the student has done or will do in the future.

General Guidelines

Different kinds of tests are appropriate depending upon some of the following general guidelines.

- **It is important to base the test on the specific skills or competencies that the course is promoting.** A course in family therapy, for example, might include performance tests on various aspects that are covered in the course, such as recording client data, conducting an opening interview, and leading a therapy session. Developing a performance test involves isolating particular demonstrable skills that have been taught and establishing ways in which the level of skill can be assessed for each student. For example, you might decide that the best way a student can demonstrate counseling skills, such as active listening, would be to have the student play the role of therapist in a simulated session.
- **Good performance tests specify criteria on which successful performance will be judged.** For curriculum areas in which it is possible to define mastery clearly, it is desirable to do so (e.g., “The student will be able to tread water for five minutes”). In most areas, however, effective performance is a complex blend of art and skill, and particular components are very subtle and hard to isolate. In these cases, it is often useful to try to highlight some observable characteristics and to define what would constitute adequate performance.

Example

In a test of teaching, students might be expected to demonstrate clarity, organization, discussion skills, reinforcement of student responses, and the like. Operational definitions for specific components to be evaluated may be phrased like the following excerpt from a teaching observation checklist: “Praises student contributions - The instructor acknowledges that s/he values student contributions by making some agreeable verbal response to the contributions. The instructor may say ‘That’s a good point,’ ‘Right, well done.’ or the like.” Such information is helpful to the student as well as the instructor who will be rating the performance.

- **Define the task as clearly as possible rather than simply alerting students to the fact that their performance will be observed or rated.** It is helpful to give students precise instructions on how the test will be structured, including how long they will have to complete the task, the conditions under which they will perform the task, and other factors that will allow them to anticipate and prepare for the test. If possible, set up a new testing situation by asking a student or colleague to go through a trial run before using the test with students so that unanticipated problems can be detected and eliminated.

- **It is important to give the same test or same kind of test to each student.** When possible, it is best to arrange uniform conditions surrounding a performance-testing situation. Students can be given the same materials to work with or the same task. Often, however, particularly in professional practice situations, it is difficult to control the context of a performance-testing situation. One nursing student may be evaluated while dealing with an especially troublesome patient, while another will be working with a helpful patient. In these situations, documenting and allowing for the contextual influences on the performance is an extremely important part of the evaluation.

In summary, the effectiveness of performance testing is directly related to how appropriate the test is, given the course objectives; how clearly the tasks are defined; how well the criteria for successful performance have been identified and conveyed; and how uniform the testing is for all students involved.



Related Chapter

Chapter 13

See **Grading** for a discussion of grading students in a performance situation.

- **Use a variety of item types.**

It is often advantageous to include a mix of item types (multiple choice, true/false, essay) on a written exam or to mix types of exams (a performance component with a written component). Weaknesses connected with one kind of item or component or in students' test taking skills will be minimized. If a mix of item types is used on one exam, items of the same type should be grouped together.

- **Be cautious about test banks.**

You should be cautious about using tests written by others. Items developed by a previous instructor or by a textbook publisher can save a lot of time, but they should be checked for accuracy and appropriateness for the given course, and whether they are written according to the standards of test construction.

- **Test early.**

You will find it helpful to test early in the semester and, if results are poor, consider discounting the first test. Students often need a practice test to understand the format each instructor uses and to anticipate the best way to prepare for and take particular tests.

- **Test frequently.**

Frequent testing helps students avoid getting behind, provides you with multiple sources of information to use in computing the final course grade (thus minimizing the effect of "bad days"), and gives students regular feedback.



Suggestion from an instructor in Information Management Systems, College of Business - “I give quizzes every week. (I don’t count two of the quizzes, giving the students a chance for a cut and to drop their lowest grade. Some of the quizzes do not require study or effort but they do tell me if the students are understanding very general concepts....) The students studied for the quizzes, and I believe they did better on the mid-term and final exam than students in other sections of the course largely because of the quizzes.”

- **Test in proportion to the emphasis a topic was given in class.**

It is important to test various topics in proportion to the emphasis they have been given in class. Students expect this practice and study with this expectation.

- **Show items to colleagues before printing the test.**

Written exams should be proofread with care and, when possible, a second person should be asked to proofread them. Tiny mistakes, such as mis-numbering the responses, can cause problems later. Also, check carefully for missing pages after collating.

- **Reuse effective test items.**

If enough test items are developed and kept out of circulation between tests, it is possible to develop a test item bank from which items that are known to be effective can

be reused on multiple versions or offerings of a test. (See Using Item Analysis to Test the Test for information on how to determine the effectiveness of test items.)

- **Do not use a series of questions in which answering successfully depends on knowing the correct answer to a previous item.**

Generally, on either a written or performance test, it is wise to avoid having separate items or tasks depend upon answers or skills required in previous items or tasks. A student’s initial mistake will be perpetuated over the succeeding items or tasks, penalizing the student repeatedly for one error.

- **Pilot-test the exam.**

A good way to detect test errors in advance is by pilot-testing the exam. You can take the test yourself or ask colleagues and/or former students to critique it.

- **Be aware of the needs of special students.**

It is important to anticipate special considerations that learning-disabled students or non-native speakers may need. You must decide whether or not these students will be allowed the use of dictionaries, extra time, separate testing sites, or other special conditions.

- **Bring extra copies of the test to class.**

Having too few copies of a written exam can be a disaster. You can avoid problems by bringing more copies of the exam than you think will be needed. Also, when duplicating the test, be certain that no pages are missing. Missing pages can pose a serious problem unless a contingency has been planned.

- **Do not interrupt students while they are taking the exam.**

Before the exam, students can be informed that they should check the board periodically for instructions or corrections. You can minimize interruptions during the exam by writing on the board any instructions or corrections that need to be made after the exam has begun and then calling students' attention to them.

Helping Students Learn from Tests

Testing's most important function is to serve as an educational tool, not simply as a basis for grading. Not only do tests direct students' studying, but also provide important corrective feedback for the student.

Returning Test Papers

Returning test papers promptly is appreciated by students and conforms to traditional learning principles. However, if you do not plan to discuss the papers, do not hand them back at the beginning of the hour or you risk losing students' attention for the rest of the hour.

Although students appreciate your returning examinations to them, there may be some question as to whether you should return multiple-choice examinations. Multiple-choice items are difficult to construct, and you may not want the items to "get out." However, you can return separate answer sheets so that your marking and arithmetic can be checked. Allow students to have copies of the examination while you go through

the test. If you follow this method, however, certain questions arise. Does such a procedure destroy the validity of the items in future tests? Do the students benefit from an exam review? These are experimental questions to which we have only partial answers, but evidence suggests that validity is not lost and that students do learn from their corrected papers, even when they do not get to keep them. Although you may not wish to spend class time quibbling over some individual items, you should make known your willingness to discuss the test individually with students.

Providing Feedback for Essays and Short-Answer Tests

The comments written on essays and short-answer tests are far more important than the grade. What kinds of comments are helpful? Look for problems that arise from a lack of ability to see relationships, implications, or applications of material. Help students find alternative ways of looking at the problem rather than simply noting that something is wrong.

Comments that provide correction and guidance may not achieve their purpose if students become so discouraged that they give up. The motivational as well as the cognitive aspects of comments need to be considered. Misconceptions must be identified, but not in overwhelming number. Encouragement and guidance for improvement should set the overall tone.



Suggestion

When you review an essay or short-answer test in class, describe what you had expected in a “good” or “satisfactory” answer, and then discuss common inadequacies. Read an example of a good answer (without identifying the student), and construct a synthetic “poor” answer as a contrast.

Reviewing Limited-Choice Tests

A small-group exercise is a technique for helping students learn from mistakes while reducing their tendency to complain about the appropriateness or fairness of test items. Instructors using this technique break the class into small groups of five to eight students. Each group discusses the test for part of the class period. When they have finished, unresolved questions are referred to the instructor as the expert. This method seems to permit dissipation of the aggressions aroused and to limit arguments to points where there are several aggrieved students.

Dealing with Special Problems

What about the student who comes to your office in great anger or with a desperate appeal for sympathy but with no educationally valid reason for changing the test grade? First, listen. Engaging in a debate will simply prolong the

unpleasantness. If you decide not to change the grade once you have heard the student out, try to convert the discussion from one of resistance to problem solving. Try to help the student find alternative modes of study that will produce better results. (“What can we do to help you do better next time?”) Encourage the student to shift from blaming you or the test toward motivation to work more effectively.



Suggestion

A technique that will reduce the number of students coming to your office in a state of high emotion is to ask students who have complaints about grades to write a paragraph describing their complaint or point of view. State your willingness to go over the test with anyone who brings in such a paragraph. This technique has a calming effect, resulting in fewer unfounded complaints and more rational discussion with those who do come to your office.

While these suggestions may save you some bitter moments, they cannot substitute for the time (and it takes lots) devoted to the construction of good tests.

Adapted with permission from: A Guidebook for University of Michigan Teaching Assistants. Center for Research on Learning and Teaching, University of Michigan and from: Teaching Tips: A Guidebook for the Beginning College Teacher (9th ed.) by W. J. McKeachie (1994).

Using Item Analysis to Test Your Test

After a test has been administered, a good way to judge its quality, particularly in the case of a limited-choice test, is to perform an item analysis. It is especially important to do this when test items will be reused or when there is sufficient doubt about students' test results to consider dropping some items as invalid when computing the final grade. Machine scanned test forms or software can be used for an item analysis. It is possible to perform an item analysis without a computer, especially if the test is short and the class size is small.

Procedures for Computing Difficulty and Discrimination Indices

- Score each test by marking correct answers and putting the total number of correct answers on the test.
 - Sort the papers in numerical order (highest to lowest) according to the total score.
 - Determine the upper, middle, and lower groups. One way to do this is to call the top 27% (some people use the top third) of the papers the "upper group," the bottom 27% (some people use the bottom third), the "lower group," and the remaining papers, the "middle group."
 - Summarize the number correct and number wrong for each group.
- Calculate the difficulty index for each item by adding the number of students from all groups who chose the correct response and dividing that sum by the total number of students who took the test. The difficulty index will range from 0 to 1, with a difficult item being indicated by an index of less than .50 and an easy item being indicated by an index of over .80.
 - Calculate the discrimination index by first calculating for both the upper and lower group students the percentage of students who answered each item correctly. Subtract the percentage of lower group students from the percentage of upper group students to get the index. The index will range from -1 to +1, with discrimination over .3 being desirable and a negative index indicating a possibly flawed item.

Table 2 illustrates **item analysis** for a simple set of scores for 37 students on a 10-item test. The names of the 10 students (approximately 27% of the total students) with the highest scores are listed as the upper group; the 10 students with the lowest scores (again, approximately 27%) are listed as the lower group; and the remaining 17 are listed as the middle group. On item 1, for example, the difficulty index was calculated by totaling the correct responses and dividing by the number of students ($19/37 = .51$). The item appears to be on the difficult end of the range.

The **discrimination index** for the same item was obtained by first calculating the percent correct for both the upper and lower groups - 20% and 90% respectively - then subtracting the percentage for the lower group from that of the upper group (.20 - .90 = -.70). This negative discrimination index indicates that the item is probably flawed. Note that the students who scored poorly on the exam as a whole did well on this item and the students who got the top total scores on the exam did poorly - the reverse of what one would expect. A mistake in the answer key or some error in the question that only the more discriminating students would catch might be the cause. If the answer key is correct, this item should be dropped from the test. Such items should be revised before being used on a test again.

Sample Test Grid for 10 Items

Item Numbers

	1	2	3	4	5	6	7	8	9	10
Upper Group										
Ellen	C	C	C	C	C	C	C	C	C	C
John	C	C	C	C	C	C	C	C	C	C
Albert	W	C	C	C	C	C	C	C	C	C
Joanne	W	W	C	C	C	C	C	C	C	C
Maria	W	C	C	C	C	C	C	C	C	C
Anne	W	C	C	C	C	C	C	C	C	C
Doris	W	C	C	C	C	C	C	C	C	C
Joshua	W	C	C	C	C	C	C	C	C	C
Barbara	W	C	C	C	C	C	C	C	C	C
Michael	W	C	C	C	W	C	C	C	C	C
# Correct	2	9	10	10	9	10	10	10	10	10
# Wrong	8	1	0	0	1	0	0	0	0	0
Middle Group										
# Correct	8	12	12	13	12	13	11	11	12	12
# Wrong	9	5	5	4	5	4	6	6	5	5

Lower Group										
Lucille	C	C	C	C	W	C	W	C	W	C
Joseph	C	C	C	C	W	C	W	C	C	C
Charles	W	W	C	C	C	C	W	C	C	W
Leslie	C	C	C	C	C	C	W	C	C	W
Jerome	C	C	C	C	C	C	W	C	C	C
Nancy	C	C	C	C	C	C	W	W	C	W
Judith	C	C	W	C	C	C	W	W	W	W
Ralph	C	W	W	W	C	C	C	W	W	W
Beth	C	C	W	W	W	W	W	W	W	C
Donald	C	W	C	C	W	C	W	W	W	C
# Correct	9	7	7	8	6	9	1	5	5	5
# Wrong	1	3	3	2	4	1	9	5	5	5
Difficulty										
Index	.51	.76	.78	.84	.73	.86	.59	.70	.73	.73
Discrimination										
Index	-.7	.2	.3	.2	.3	.1	.9	.5	.5	.5
	C =	Correct								
	W =	Wrong								

Adapted with permission from: Teaching at the Ohio State University: A Handbook. Center for Teaching Excellence (1990).

Cheating on Tests

The University has an Academic Honor Code that calls for the coordinated efforts of faculty members and students to uphold academic integrity and combat academic dishonesty, including cheating and plagiarism. The Academic Honor Code includes descriptions of violations of the code, statements of student and faculty responsibilities for upholding the code, and explanations of academic penalties for violating the code. A description and information of the Academic Honor Code can be found in the current **Student Handbook**.

Preventing Cheating

- **Reduce the pressure.**

The first action you can take is to reduce the pressure on students. While you cannot influence the general academic atmosphere that places heavy emphasis on grades, you can influence the pressure in your own course. One method to accomplish this is to provide students with several opportunities to demonstrate their achievement of course objectives rather than relying upon a single examination.

- **Make reasonable demands.**

A second way to reduce cheating is to make sensible demands. Write fair tests and design reasonable assignments. Some cheating is simply the result of frustration and desperation arising from assignments that are too long to be completed adequately or tests that require the memorization of trivial information. Remember, some students view cheating as a way of getting back at an unreasonable instructor.

- **Treat students as individuals.**

Work to develop and maintain students' sense of individuality and having a personal relationship with you and their classmates. Students are not as likely to cheat in situations where they are known as individuals, whereas they may be tempted to cheat in situations where they feel they are anonymous members of a crowd. If a large course has regular meetings in small discussion or laboratory sections, there is likely to be less cheating if the test is administered in these groups than if the test is administered en masse. Moreover, if it is in their regular classroom, they will perform better.

- **Show an interest in your students.**

Cheating is more likely to occur when students think the instructor is disinterested and unconcerned. Instructors often feel that any show of active proctoring will indicate that they do not trust the students. However, it is possible to convey a sense of alert helpfulness while walking between desks and watching for questions.

- **Use alternate seating.**

The most common form of cheating is copying from another student's paper. To minimize opportunities for copying, try to recruit proctors and administer exams in a room that is large enough to enable students to sit in alternate seats. Before students arrive, write on the chalkboard: "Please sit in alternate seats."

- **Use alternate test forms.**

Another way to reduce cheating is to use two or more alternative forms of the test. This method can be achieved by simply scrambling the order of the test items. Instructors who want the test items to follow the order that the material was covered in the course can scramble the items within topic areas.

- **Be careful with extra copies.**

Do not leave copies of tests lying around your office, the typist's office, or photocopy room.

Adapted from: *Teaching Tips: A Guidebook for the Beginning College Teacher* (9th ed.), by W. J. McKeachie, Lexington, MA: D. C. Heath. (1994).

Handling Cheating

Despite preventive measures, almost every instructor must at some time or other face the problem of what to do about a student who is cheating. Policies for handling cheating are set by the University as well as by departments. FSU's Faculty Handbook provides specific information about university policy. There are assessment devices, other than tests, that can be used to provide measures of student performance, including:

- Essays
- Term papers
- Research reviews
- Reports
- Case studies
- Portfolios
- Projects
- Performances
- Peer evaluation
- Mastery
- Simulations

Just as with tests, the underlying principles to keep in mind as you introduce alternative assessment tools are validity and reliability. A tool you use for measurement will be valid as long as it measures student learning of goals and objectives set for the course. The measurement will be reliable if you expect to get similar results administering the chosen assessment tool to the same group of people again. It will have the additional benefit of reusability, if it can be used in multiple instances, including future classes.

Available media and equipment influence the choice of assessment tools. But the principle factor involved in the choice of an assessment strategy is the overall design of the course. Good design principles demand that the assessment strategies be chosen as part of the overall instructional plan before the course actually starts.

- **Essays** are written assignments in which the student is the source of information. Essays report on things the student knows or thinks. Reference is not a major part of an essay. But an essay requires the student to use high-level thinking skills. There may be preparatory activities involved with writing an essay. For instance, an assignment may ask a student to read several articles from several viewpoints and then derive his own viewpoint from the articles. The expectation in giving the assignment is that the student will apply reasoning skills and reach a conclusion that is well-reasoned. The actual position reached is not the main value of the essay, and should not be evaluated unless your objective is to have students' opinions agree with your own.

Essays expose student reasoning-processes. An assignment of a practice essay (that is not figured into the course grade) near the beginning of a course gives the instructor an idea of whether the student's reasoning skills are adequate for pursuing the course. If a change in reasoning skills is a desired course outcome, an essay assigned near the end of the course is a good way to tell whether the desired skills have been attained.

- **Reports** (and term papers) usually have a specific topic that may be assigned by the instructor or selected by the student. When a student reports on a set of facts or events, accuracy of the student's description is the main concern. The report often includes a provision for commentary by the student. The student's commentary is presumed to reflect the student's point of view accu-

rately about the facts, events, or issues of the report. Research on which a report is based may use a variety of sources including experimentation and documentation. The amount of research for a report varies. In a report based on documentation, credit for quotations and concepts should be included.

- **Research reviews** ask a student to find out what research about a topic area has been done. Unless the student is asked to synthesize the results of the research, reviews offer little room for a student's free expression or creativity. If the same research review is assigned to a group of students or a class, duplication of the research found should be anticipated. The assignment measures the ability of the student to use available research tools and the ability to judge whether articles found qualify as appropriate references for the subject at hand.
- **Case studies** are often associated with problem-based learning. They are used to assess a learner's ability to analyze, make decisions, and solve problems.
-



Related Chapter

Chapter 8

Using Active Learning in the Classroom, provides learning through case studies is addressed.

Case studies measure depth of learning to a greater extent than most limited choice tests, which focus on memorization skills. Used as an assessment tool, the

instructor usually creates the case that can be contained within the allocated assessment time. Like the case studies used for learning, case-study assessments contain a number of circumstantial descriptions that provide guidance through the project. While there are self-consistent solutions, there are no "perfectly right" answers. Students look at the circumstances, bring them into their own personalized conceptual frameworks, and then try to provide solutions. Consequently, answers may be phrased in various ways, but must include all salient points and exclude inaccurate points.

Advantages: Case studies assess the readiness of learners to use their skills in real world contexts. The student may expose both the process of working through a case and the results obtained. This allows the assessment to focus on either the process or the product. The student should be informed ahead of time whether process or product will receive the greater weight. Cases are equally easy to present in computer-based, hard-copy, or audio-based form. They lend themselves to either collaborative efforts or individual efforts.

Student answers to case-type problems may, with permission of the students involved, become models for answers to cases delivered in subsequent classes. If prior student work is used to model poor performance, care must be taken to keep the student's name and identity anonymous. It is probably better to con-

struct poor examples, if they are needed, rather than use student work.

Limitations: The cases may require considerable effort to create or select. The tendency to collaboration makes it difficult for teachers to assess individual contributions to team efforts. Grading case studies take a long time and tends to be highly subjective. Subjectivity is greatly reduced, however, by using guidelines and rubrics that specify the features that will be assessed in the answer.

Contextual considerations: Case studies can be used with almost any available media. Cases can be presented in print, audio, video, computer, or Internet depending on what the learners have available. Proposed solutions can be designed for presentation using any of these media choices. Where the Internet is available, using collaborative learning tools, such as those included in Blackboard, allows tracking individual participation in collaborative activities, provided all parties have agreed to use these tools. This means that where cases are assigned to groups, individual contributions to group effort can be assessed in a limited sense. A tracking system may note the number of contributions an individual makes to a discussion, but rarely evaluates the quality of those contributions. Checking samples from an archive of the discussion often gives a truer picture of individual student efforts.

- **Portfolios** contain all the work that students collect to demonstrate how their work has progressed and developed over time. The student, over the length of

the course, will have been involved with and probably completed several projects. In most cases, it is wise for teachers to set standards for portfolio contents in terms of number, quality, and size of projects to be included. Fine Arts and Performing Arts classes are likely to include portfolios among the assessment tools used. Other instances where tangible products result from student efforts are also good places to use portfolios. For instance, a business class may create a portfolio that includes a business plan, some presentations, and an implementation timeline as evidence of completion of course objectives.

Advantages: Assessing portfolios encourages extended thinking and reasoning. Where privacy is an issue, sharing project contents should be strictly controlled. The student is completely responsible for decisions about what to include and is therefore led to considerations of quality in what will be submitted. Except for setting standards, teachers do not have much to do with the construction of the portfolio and are free to guide learners toward completion of their portfolio projects. A wide variety of media choices is available for the types of projects that may be included.

Portfolios are useful in a wide range of subject areas. Besides the arts, where it makes sense to let students demonstrate their capability with produced work, such areas as writing and even mathematics have been assessed with portfolio type assessment tools. In writing, student style and organizational skills are demonstrated with portfolios. In mathematics, data

analysis and problem solving may also be assessed with portfolio work.

Limitations: Grading is highly subjective. The teacher sees only the successes of the student and leaves all failed projects out of grading considerations. Grading portfolios can take a long time. Portfolios can show development on a broad spectrum of work or allow students to concentrate extensively on very narrow areas of a field of study. This strength of the portfolio approach allows students to mask weaknesses in other areas. Extensive use of portfolios may encourage students to repeat earlier successes rather than to undertake new challenges.

Contextual considerations: Projects must be constructed in a media form compatible with the available media. Mediation of the projects in the portfolio to match delivery system capabilities may distort the value of the project. For example, a portfolio of art work loses scale information and impact when displayed on a computer monitor. Where such distortions are likely to happen, the student should be made aware of them and requested to take measures to offset the distortions (through labeling, etc.). Also, handing a portfolio physically puts a large and potentially valuable body of student work at risk.

- Term papers are valuable to students because they provide them with an opportunity to be experts in small but relevant areas of the field. They should be limited to one

per term. Term papers are long for most students and are a significant part of the work for the term. A term paper should be introduced early in a course and collected near the course's end. It should contain some requirement for research, and a strong indication that the student has mastered the course material as it was presented over the term of the course. A term paper is a type of project, and the characteristics and recommendations for projects apply (see next bullet).

Suggestions for making term papers more effective measures of learning:

- Ask students to write to readers other than you, such as peers, experts in the field, or specific journals.
- Clarify what the final term paper should do: classify, explain, summarize, demonstrate, generate, or design.
- Let students know your expectations concerning:
 - ◇ Academic discourse conventions
 - ◇ Level of formality
 - ◇ Structure: introductions, bodies, conclusions, and internal organization options
 - ◇ Formatting instructions: length, margins, typing, cover page, page numbering, and documentation style. Give samples, if possible.
 - ◇ Charts, graphics
- Assist in the writing process:
 - ◇ Students bring in drafts and respond to each

- other's work.
- ◇ Conduct individual conferences.
 - ◇ Photocopy a past student's draft and critique it as a class.
 - ◇ Encourage students to take their drafts to the Writing Center.
 - ◇ Schedule the workload.
- When grading term papers:
 - ◇ Avoid being too directive, commenting on every grammatical error or global problem, and avoid making vague or generic comments.
 - ◇ Respond to strengths and weaknesses.
 - When responding, save yourself time by:
 - ◇ Marking patterns in grammatical errors, or have students find the errors.
 - ◇ Focusing on three or four major issues.
 - ◇ Having students peer review the term papers before turning them in.
 - ◇ Having students visit the Writing Center.
 - ◇ Designing your assignment carefully.
 - ◇ Using a grading rubric.
 - Projects - A teacher may request one large project or several smaller projects during a student's course of study. The assessments of student performance on the

project(s) collected may make up the whole or a part of a student's course performance. Monitoring the student's ongoing progress toward completion of a project moves the emphasis of instruction and assessment away from outcome and toward process. Large projects create an opportunity for the instructor and the student to work with each other. Typically, an instructor will assign a large project to the student, then check the student's progress at various stages, offering advice for changes to be made along the way.

Advantages: Students doing projects work outside the boundaries of the classroom. Classroom time-constraints play a limited role in how projects turn out. Student effort is given to organizational activities along with problem solving. This allows the instructor to see the student's work at its pinnacle, much like a portfolio. Presentation of some or all of the projects by their creators can be a classroom enhancement activity. Students should know in advance whether all projects or a sample of projects will be presented. The student-to-teacher interactivity involved in most projects provides feedback to learners at the most appropriate time in a student's learning process - while a project is being done. Interaction between teacher and student helps keep the student on schedule toward project completion.

Limitations: Since so much of the project is done outside the classroom setting, it is very difficult to monitor what the student is doing while completing

the project(s). Different levels of acceptable outside help may need to be defined. The pacing of projects comes from both the student and the instructor, and consequently does not faithfully represent the student's own ability to set a pace. Providing feedback to the student in a timely manner requires large amounts of teacher time and frequent teacher check-ins.

Contextual considerations: Interactive projects are very well suited to Internet-based teaching tools. Tools for project interactivity over the Internet, such as "Common Space" (for writing), are becoming readily available at low cost. The use of these complex tools greatly facilitates working with this strategy, although there is a need to learn how to use the tools that may take away from available contact time. Feedback on projects must be timely, since pacing can be easily upset by long teacher delays.

- **Performances** require a student to perform in a classroom or at a special facility where performances of the type in question are done. Typically, there will be some measurement of the quality of the performance requested. The instructor alone, the instructor and the rest of the class, or a larger invited group, may view performances.

Advantages: Pre-established criteria help the student by clarifying the evaluation standards, letting students see strengths and weaknesses that they have

relative to these criteria, and demonstrating that the grading system used is consistent.

Limitations: The number of available criteria that can be applied to measurement of student performance is incredibly high. No single checklist is likely to encompass every aspect of what students do when they are asked to perform. The richness of available criteria, however, does not make assessment of a live performance impossible. Instead, it makes precision in matching the choice of included criteria to learning objectives more critical than ever. For example, if you were making a performance checklist for a dance performance, you would almost certainly include some criteria related to the execution of the performance. But would you include criteria related to the choreography? You would probably only include this if the student was responsible for the choreography. Would you include a measure for the warmth of a student's smile while dancing? It depends on whether your instruction included any encouragement of the act of smiling as something that enhances stage presence.

You want to be fair to the students as much as possible in the checklist criteria you include. If your learning objectives are at particular levels, the performances on your checklist should be at the same levels. Creating scales and rubrics provides guidance for measuring what are often spontaneous and subjective performances. In general, students should be aware of the objective criteria on which performances

are measured.

- **Peer Evaluations:** The idea behind the use of peer evaluation as an assessment tool is that a student's peers, who have had to work with and contend with a student, have a good idea of that student's contribution level. The higher the grade levels of the students, the more likely this is to be the case. By asking students to review other students' products, results, or performances, you can take account of experiences in which you were not directly involved. Peer evaluation often involves the use of a measurement instrument distributed multiple times. This presumes that all of a student's peers participate in the evaluation. Sometimes, the number of instruments to be completed and counted may be reduced through varying both the assignment of partners and the assignment which peers get to review. Some formula for compiling the instrument results yields an indication of each student's peer evaluation score. Peer evaluation, when used at all, should be only a part rather than the whole of a student's final grade.

Advantages: Peer evaluation fills a gap in the usual assessment process that exists because so much of a student's performance is unseen. Peer evaluation instruments, while subjective on an individual basis, provide data through which a quantitative measure of subjective judgments is accumulated. The feedback of measured information to the student, in carefully chosen circumstances, may motivate improvements in student performance. Group dynamics are being measured with these instruments, which is important

in environments that value active participation in collaborative activities.

Limitations: Peer evaluation may measure student popularity or some phenomenon other than the one the instructor wants to assess. Although the instruments are used to gather data, the data are an accumulation of subjective judgments. Summary of the measurement-instrument results is time consuming. A set of standards should be provided and explained to students, or students may agree among themselves on the standards they will use in determining peer performance.

Contextual considerations: In computer-mediated, distance-education environments, collaboration is often a significant part of the learning experience. Peer evaluation is the only tool that measures this collaboration from the group's point of view. The negative influences of poor in-group performances by some students may be brought to light. Where computers are used to mediate communication in the educational environment, they can be used to aid in gathering and summarizing peer evaluation data making peer evaluation strategies easier to use.

- **Mastery Models** - When it is important that a skill be mastered, an "all or nothing" approach, similar to pass/fail, may be the best indicator. This is particularly true when a skill that will see later use in the learning process

is first being learned. To assess using a mastery model, it is typical to assign a project that involves the use of the new skill, so that handing in a successfully completed project serves as an indication of the skill having been mastered. In many cases, the learner may take an unlimited number of tries without penalty, but will pass once mastery has been demonstrated. In essence, this model is like pass/fail for the steps along the way to achieving course objectives.

Advantages: Mastery models are not competitive - everyone can and will master the skill. Such models have high validity and reliability, and they provide a clear and direct measure of success in reaching learning objectives. Students also have the ability to avoid the need to relearn redundant material.

Limitations: Mastery models are only applicable in skill-learning situations. While they measure the mastery of a skill, there are different levels of mastery that are not measured beyond the minimum competency level. Because everyone passes eventually, mastery leaves open the question of how to give grades. When everyone succeeds, there needs to be a difference between A and D, but this method incorporates no distribution of grades that may be used for determining the difference. While applying the principles of mastery learning helps students get through courses, the non-competitive nature of the learning makes it difficult to assess inside a competitive framework.

- **Simulations** - In an assessment that uses simulation,

students are placed into an environment that, in many significant ways, looks and behaves like the environment where learning will actually be applied. They are given opportunities to perform in the simulated environment. Some record of their performance is used as the basis for assessment.

Advantages: The use of simulations reduces the student's exposure to situations that could have strong negative consequences if performance was done improperly in a real-world situation. A classic case of simulated assessment is the airplane cockpit simulator that monitors student performance in handling the controls of an airplane but will not crash. Simulations do not have to reach this level of complexity, however. Students have learned social skills such as job interviewing by playing their roles as if they were really going through an interview process.

The simulation provides a more accurate measure of performance than just asking students to describe how they would do something. Simulations can be repeated to achieve continuous improvement until a standard level of performance is reached.

Limitations: Simulations must be designed for each situation in which performance is being assessed. Designing and building the simulations is costly and time consuming. Once a student is made aware that the situation is a simulation, stresses associated with real world performance are significantly reduced, resulting in an inaccurate measure of the student's actual capacity to perform.

Contextual considerations: The simulation exercise must be tested and calibrated at the student's location. Many simulations are done with computers, and this makes their assessment results easy to pass on to other computers.

Resources on Testing

Books/Articles

- Anderson, P. S. (1987). *The MDT innovation: Machine scoring of fill-in-the-blank tests*. (ERIC Document Reproduction Service No. ED 307 287)
- Astin, A. W. (1991). *Assessment for excellence: The philosophy and practice of assessment and evaluation in higher education*. New York: American Council on Education/Oryx Press.
- Ben-Chiam, D., & Zoller, U. (1997). Examination-type preferences of secondary school students and their teachers in the science disciplines. *Instructional Science*, 25, (5), 347-67.
- Bloom, B. S., & Madaus, G. (1981). *Evaluation to improve learning*. New York: McGraw-Hill.
- Boaler, J. (1998). Alternative approaches to teaching, learning and assessing mathematics. *Evaluation and Program Planning*, 21 (2), 129-141
- Cashin, W. E. (1987). *Improving essay tests*. (Idea Paper No. 17). Manhattan, KS: Kansas State University, Center for Faculty Evaluation & Development.
- Clegg, V. L., & Cashin, W. E. (1986). *Improving multiple-choice test*. (Idea Paper No. 16). Manhattan, KS: Kansas State University, Center for Faculty Evaluation & Development.
- Cooke, J. C., Drennan, J. D., & Drennan, P. (1997). Peer evaluation as a real life-learning tool. *The Technology Teacher*, 23-27
- Cross, K. P., & Angelo, T.A. (1993). *Classroom assessment techniques: A handbook for college teachers* (2nd ed.). San Francisco: Jossey-Bass.
- Duffy, T. M., & Cunningham, D. J. (1996). Constructivism: Implications for the design and delivery of instruction. In D. H. Jonassen (Ed.), *Handbook of research for educational communications and technology* (pp.170-195). New York: Lawrence Erlbaum Associates.
- Erwin, T. D. (1991). *Assessing student learning and development: A guide to the principles, goals and methods of determining college outcomes*. San Francisco: Jossey-Bass.
- *GLE: Grade Level Examination. Ensuring Academic Success* (1991). San Diego, CA: Tudor Publishing. (ERIC Document Reproduction Service No. ED 363620)
- Hansen, J. D., & Dexter, L. (1997). Quality multiple-choice test questions: Item-writing guidelines and an analysis of auditing test banks. *Journal of Education for Business*, 73 (2), 94-97.
- Jacobs, L., & Chase, C. (1992). *Developing and using tests effectively: A guide for faculty*. San Francisco: Jossey-Bass.

- McKeachie, W. J. (1994). Tests and examinations. In W. J. McKeachie (Ed.), *Teaching tips: Strategies, research, and theory for college and university teachers* (9th ed., pp.71-93). Lexington, MA: D.C. Heath.
- LaPierre, S. D. (1992). *Mastery-level measurement: An alternative to norm-referenced intelligence testing*. Reston, VA: National Art Education Association. (ERIC Document Reproduction Service No. ED 346 024)
- Mager, R. F. (1997). *Measuring instructional results* (3rd ed.). Atlanta, GA: Center for Effective Performance.
- Mehrens, W. A., & Lehmann, I. J. (1991). *Measurement and evaluation in education and psychology* (4th ed.). New York: Holt, Rinehart & Winston.
- Metzger, R. L., Boschee, P. F., Haugen, T., & Schnobrich, B. L. (1979). The classroom as learning context: Changing rooms affects performance. *Journal of Educational Psychology*, 71, 440-442.
- Miller, H. G., Williams, R. G., & Haladyna, T. M. (1978). *Beyond facts: Objective ways to measure thinking*. Englewood Cliffs, NJ: Educational Technology Publications.
- Milton, O. (1978). *On college teaching: A guide to contemporary practices*. San Francisco: Jossey-Bass.
- Myerberg, N. J. (1996). *Performance on different test types by racial/ethnic group and gender*. (ERIC Document Reproduction Service No. ED 400 290)
- Natal, D. (1998). *Online Assessment: What, Why, How?* (ERIC Document Reproduction Service No. ED 419 552)
- Newmann, F. M., & Archbald, D. A. (1992). The nature of authentic academic achievement. In H. Berlak, T. Burgess, J. Raven, & T. Romberg (Eds.), *Toward a new science of educational testing and assessment* (pp. 71-83). Albany, NY: State University of New York Press.
- Nitko, A. J. (1983). Item analysis: Using information from pupils to improve the quality of items. In A.J. Nitko (Ed.), *Educational tests and measurement: An introduction* (pp. 284-301). New York: Harcourt Brace Jovanovich.
- Ory, J. C. (1979). *Improving your test questions*. Urbana-Champaign: University of Illinois, Office of Instructional Resources.
- Ory, J., & Ryan, K. (1993). *Tips for improving testing and grading*. Newbury Park, CA: Sage.
- Recess, M. D. (1997, March). *Constructs assessed by portfolios: How do they differ from those assessed by other educational tests*. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.
- Resnick, L. B., & Resnick, D. P. (1992). Assessing the thinking curriculum: New tools for educational reform. In B. R. Gifford & M. C. O'Connor (Eds.), *Changing assessments - Alternative views of aptitude, achievement and instruction* (pp. 37-75). Boston: Kluwer Academic Publishers.
- Roos, L. L., Wise, S. L., Yoes, M. E., & Rocklin, T. R. Conducting self-adapted testing using Microcat. *Educational and Psychological Measurement*, 56 (5), 821-827.

- Smith, C. R., & McBeath, R. J. (1992). Constructing Matching Test Items. In R. J. McBeath (Ed.), *Instructing and evaluating in higher education: A guidebook for planning learning outcomes* (pp. 199-223). Englewood Cliffs, NJ: Educational Technology Publications
- Straetmans, G. J. J. M., & Eggen, T. J. H. M. (1998, January-February). Computerized adaptive testing: What it is and how it works. *Educational Technology*, 45- 51.
- Svinicki, M. D. (1976). The test: Uses, construction and evaluation. *Engineering Education*, 66 (5), 408-411.
- White, E. M. (1985). *Teaching and assessing writing*. San Francisco: Jossey-Bass.
- Zaremba, S. B., & Schultz, M. T. (1993). *An analysis of traditional classroom assessment techniques and a discussion of alternative methods of assessment*. (ERIC Document Reproduction Service No. ED 365 404)

Website

- *Principles and Indicators for Student Assessment Systems*. FairTest, The National Center for Fair & Open Testing. <http://www.fairtest.org/princind.htm> (accessed November 3, 2005).