



The Scarecrow Press, Inc.
Metuchen, N.J., & London
1979

Anthony S. Jones
Lawrence W. Bagford
Edward A. Wallen



STRATEGIES FOR TEACHING

TEACHER-MADE TESTS

Most textbooks dealing with testing stress the use of tests to measure instructional effectiveness. However, most teachers use tests to furnish a basis for assigning a grade. Rather than become embroiled in a philosophical discussion—since teachers would still use tests for assigning grades and test experts would still stress tests as a measure of instructional effectiveness—let us say that tests can be used for both purposes. Also, they can be misused for either purpose.

Whatever the case, the several types of testing avenues available to the classroom teacher have inherent qualities much like the instructional strategies in this book. Rather than get bogged down in heavy statistical treatment, our aim here is to help teachers better use the testing avenues available to them. This approach does not preclude the need for teachers to become aware of statistics as applied to testing. In fact, all teachers should have formal study in properly applying statistics to the classroom situation on a scale not limited only to testing. That, however, is beyond the scope of this part of the book. Our focus is upon usage of oral, performance and written means of classroom testing as designed by the teacher. Each of these means will be examined in terms of its (1) definition and description, (2) advantages or special purposes, (3) disadvantages or limitations, and (4) guidelines for maximum utilization.

Before proceeding directly to this examination, certain testing concepts must be considered. These concepts will be used as criteria by which the testing means will be compared.

The first concept is that of *validity*. In test jargon, validity is often said to be the degree to which a test measures what it purports to measure. For our use in examining testing approaches, we are interested in what types of behaviors the various approaches best

measure. Validity is the most important of the concepts since we must be sure we are measuring what we think we are measuring.

Next comes *reliability*. Reliability deals with the consistency of results in measuring the same thing. One way of looking at reliability is to expect each of a group of students to get the same, or nearly the same, scores on a test-retest basis.

The third criterion we will deal with is *objectivity*. Objectivity really deals with the scoring of the test. If, in the scoring process, the scorer's subjective judgment does not enter in, the test is objective. In this case, the student would get the same score no matter who scored the test.

The fourth factor is *comprehensiveness*. Comprehensiveness refers to the degree of sampling of what has been taught. Obviously a teacher on a semester exam cannot test everything taught during the entire semester. However, a proportionate sampling of the major teachings can be made. The wider the range of subject matter tested and the truer the proportion tested, the more comprehensive the test is said to be.

The fifth area considered deals with *practical matters* involved in testing. These matters include (1) the time involved in constructing and scoring the tests, (2) the efficient use of class time for testing, and (3) the ease of constructing, administering, and scoring the test.

As the various testing approaches are examined, there will sometimes be a focus on test "type" and then on test "item" at other times. The reasons for this will be seen as each is examined.

COMPLETION

Definition and Description

Completion, or fill-in-the-blanks, items are those in which the student provides one or more missing words. This type of question calls for the students to recall the proper response.

Advantages or Special Purposes

1. Completion items are best used in cases where the student is expected to recall specific facts (names, dates, places, events, definitions, etc.).
2. Little opportunity for guessing is provided.
3. Completion items are relatively easy to construct.
4. This item is more comprehensive than the essay type in that it provides for greater sampling of learning in the same amount of time.

Disadvantages or Limitations

1. Overuse of the Completion test approach often creates an overemphasis on memorization on the part of the learner.
2. The higher level cognitive skills are not tested in this manner.
3. While less subjective than the essay item, the Completion item still lacks some objectivity. The teacher often will need to decide if the given answer is "close enough" or "means the same thing."

4. It is tempting for the teacher to copy statements directly from the textbook, thus creating poor items.
5. Teachers often use too many blanks, leaving a vague and ambiguous puzzle to be solved.

Guidelines for Maximum Utilization

1. Use only one blank in an item.
2. Write the item so specifically that there can be one and only one answer.
3. Place the blank at the end of the statement. If this is difficult, write the item in question form.
4. Paraphrase sentences taken from the textbook.
5. Make a model with the correct answers for use in scoring.

Questions to Promote Discussion

Below are questions which contain some errors. Using the above material, identify the errors and explain how the items should have been constructed. (Cite the specific statement in the text whenever possible.)

Completion Items (Not necessarily parts of the same test)

1. A. Lincoln was born in_____.
2. Paraphrase sentences taken from the_____.
3. _____the blanks at the end of the statement.
4. _____only one_____in_____.

ESSAY

Definition and Description

In an Essay examination, the student writes a response to one or more questions requiring an answer of at least a sentence. Such responses range from a simple sentence for an elementary school student to one requiring several days for a graduate student.

Essay examinations can be used to measure either content (the answer) or process (how well written) or both. The important point is that teachers often have not decided or have not explained to the student what "counts." Thus the student is sure to ask immediately before answering the first question, "Does spelling count?"

Advantages or Special Purposes

1. Essay questions are best used for objectives dealing with the improvement of written skills, interpretation and use of data, creative expression, organization of ideas, developing a position, and thematic development.
2. Knowing that Essay items will be used usually motivates the student to study more thoroughly.
3. While the construction of good Essay questions is more difficult than is generally believed, they are easier to construct than most types of test items.
4. "Cheating" or copying another's work is virtually impossible with proper use of essay tests.
5. The opportunity for guessing is almost non-existent.

Disadvantages or Limitations

1. Essay items are the most subjective of all of the test items. Their lack of objectivity has been demonstrated in research time after time.
2. With such a lack of objectivity, the Essay item is low in reliability.
3. Comprehensiveness of the Essay item is poor since a relatively little amount of material can be covered in a given time.
4. Large amounts of time are needed to score Essay items.
5. Essay questions are inefficient in the use of class time since the physical act of writing is much slower than the answering of objective-type items.
6. Essay items are not easy to score. They require reading time, and students have a variety of ways of expressing themselves, which demands more time for interpretation.

Guidelines for Maximum Utilization

1. Use the Essay means of evaluation only for those objectives for which it is best suited.
2. Decide whether you are concerned about content or process, or both, and explain to students.
3. State the question as *specifically* as possible so that students will know what you want.
4. Minimize the weakness of subjectivity by:
 - a. Developing a model-outline answer against which all student answers can be compared;
 - b. Scale the number of points to be awarded for each portion of the model-outline answer;
 - c. Score all papers anonymously; and,
 - d. When several specific essay questions are used, score the

first question on each paper before going on to the next question, and so on.

5. Explain the reason why each incorrect answer is not considered correct.
6. Review, in class, the model-outline answer for each item and how you scored the papers.
7. Beside each test item, list how many points each is worth, thus allowing students a basis for gauging the relative importance each question.

Questions to Promote Discussion

Below are questions which contain some errors. Using the above material, identify the errors and explain how the items should have been constructed. (Cite the specific statement in the text whenever possible.)

Essay Items (Not necessarily parts of a single test)

1. Discuss the Civil War.
2. List, in order, the first ten presidents of the U.S.
3. Compare and contrast Communism and Socialism in terms of:
 - a. the control of industry.
 - b. the ownership of property.
 - c. efficiency of agricultural practices.

MATCHING

Definition and Description

The Matching item is a set of two lists in which each item in one list is related to an item in the other list. The instructions given explain the bases on which the various items are related. For ease of identification, one column (the left) is often called the “statement” and the other (the right) the “answer.”

Advantages or Special Purposes

1. The Matching item is best used for those objectives which require the learner to associate related data.
2. Objectivity in scoring is present in this item.
3. This item is comprehensive in that large amounts of learning can be sampled in a short period of time.
4. Advantage #3 insures efficient use of class time.
5. Little time and effort is needed in scoring Matching items.

Disadvantages or Limitations

1. Often, a sufficient quantity of related data does not exist to consider using Matching items.
2. When an insufficient amount of related data exists teachers often manufacture items, thus creating poor questions.

3. Instructions regarding the basis of Matching are often difficult to write.
4. Directions for marking the correct answer are difficult to write.

Guidelines for Maximum Utilization

1. Use Matching items only where a sufficient number of data exists.
2. Items should be homogeneous. This means that all “statements” and “answers” must be dealing with the same thing (people or places or dates, etc.).
3. Explain the basis for Matching.
4. Avoid giving grammatical clues.
5. Avoid perfect matching so that students do not get the last answer by the “process of elimination,” i.e., have extra items in the answer column.
6. Keep the lists under a dozen “statements” and “answers.”
7. Arrange the “answers” in some order, generally alphabetical, to aid in finding the correct response.
8. Present clear directions regarding the manner in which the student is to indicate his/her response.

Questions to Promote Discussion

Below are questions which contain some errors. Using the above material, identify the errors and explain how the items should have been constructed. (Cite the specific statement in the text whenever possible.)

Matching

Match the following:

(Statement)

(Answer)

1. Discovered the Law of Gravity.

a. Mississippi

2. First president of the U.S.

b. 1976

3. Longest river in the U.S.

c. Newton

4. Largest city in Illinois.

d. Harvard

5. First college in the U.S.

e. Chicago

6. Bicentennial.

f. Washington

MULTIPLE-CHOICE

Definition and Description

With the Multiple-Choice item, the learner selects from among two or more alternatives the one which correctly completes the statement or answers the question. The incomplete statement or question presented is called the stem.

This item type seems to be held in higher esteem by tests and measurements people than by classroom teachers. Technically, the Multiple-Choice item is a superior item. However, it must be recognized that the item has some weaknesses which bother classroom teachers.

Advantages or Special Purposes

1. The Multiple-Choice item can be used to measure a wide variety of objectives.
2. Scoring can be done on an objective basis.
3. Reliability is generally good.
4. These items permit a fairly wide sampling of learning in a short period of time; thus they are high in comprehensiveness and efficient use of class time.
5. Multiple-Choice items are easily and quickly scored.

Disadvantages or Limitations

1. Multiple-Choice items are the most difficult to construct.
2. In constructing these items, much time is required.
3. Because test "experts" extoll the virtues of Multiple-Choice items, teachers often "force" material or objectives which could be better measured other ways into the Multiple-Choice format.
4. Multiple-Choice items are not entirely free from "guessability."

Guidelines for Maximum Utilization

1. Use Multiple-Choice items to measure only those objectives for which they are compatible.
2. Use at least four alternatives for each statement or question.
3. Write in question form.
4. Insure that each alternative is plausible enough to be at least considered by the student.
5. Avoid grammatical "hints" in the stem as to the correct answer (singular-plural, etc.).
6. Avoid the use of the word "not" in the stem. If "not" must be used, underline the word so the student doesn't perceive it is a "trick" question.
7. Avoid any pattern of answer position. Some teachers tend to make the last alternative the correct one.
8. Make all alternatives of near-equal length.
9. If you use both "all of the above" and "none of the above" in the same item, do so in that order.

10. Avoid using either “all of the above” and “none of the above” *only* when it is the correct answer. Use it at times when it is not the correct answer.

Questions to Promote Discussion

Below are questions which contain some errors. Using the above material, identify the errors and explain how the items should have been constructed. (Cite the specific statement in the text whenever possible.)

Multiple-Choice (Not necessarily items of a single test)

1. How many states are there in the United States?
 - a. 50
 - b. 48

2. A unit of weight is an
 - a. mile
 - b. foot
 - c. ounce
 - d. minute

3. The primary colors are
 - a. red
 - b. yellow
 - c. blue
 - d. none of the above
 - e. all of the above

ORAL TESTING

Definition and Description

Oral examinations in which a group of established experts attempt to assess the knowledge of an individual neophyte are commonplace in the graduate schools of American universities. Below that level, however, the use of oral examinations is usually limited to the teacher asking questions of individual students in the classroom setting.

Basically, an oral examination can be concerned with either content or process. By "content" is meant that the substance of the student's answer is of prime import rather than how well the answer was said. "Process," on the other hand, refers to the manner in which the student answered the question. The concern here is for the development of oral skills.

Advantages or Special Purposes

1. The oral approach can be used for those objectives which require the learner to explain, tell, name, etc.
2. This approach is the only one which can be used to assess the learner's development of oral skills.
3. Only an average amount of time is required to construct questions.
4. Some students are psyched out by written tests and perform with more validity on an oral examination.

Disadvantages or Limitations

1. The oral approach to testing lacks objectivity in scoring. This is why teams of judges are used in speech contests and debates. The subjectivity of the scorer enters into the scoring.
2. Since this means of testing lacks objectivity, it also lacks reliability.
3. Comprehensiveness is low since the teacher can ask only a limited number of questions in a given period of time, thus limiting the learning sampled.
4. Class time is not efficiently used since the other students are uninvolved while one student is participating.
5. Good questions are difficult to design. Too often teachers do this "off the top of their heads," the result being vague and unclear questions.
6. Students often give poor answers because they lack oral skills, not because they don't know the information.
7. Giving incorrect responses or not being able to respond in front of their peers often impairs students' belief in themselves.
8. Teachers often are unsure of whether they are measuring content or process, or both.
9. Scoring is difficult because not all students have to deal with the same questions.

Guidelines for Maximum Utilization

1. Use Oral Testing basically to measure the development of oral skills.
2. If you must use Oral Testing for other purposes, do so sparingly and follow the remaining guidelines.

3. Decide whether you are measuring content or process. If you are not specifically measuring oral skill development, focus on content only.
4. Write out your questions in advance, being sure they are clear and unambiguous. (Refer to the *questioning* strategy for help on this.)
5. Make sure the questions proportionately sample the content taught.
6. Ask the question, pause, then call on a student. This will encourage all students to think about the answer.
7. Avoid a pattern of calling on people, alphabetically, down the row, etc. A pattern known by the students means one only has to think when his/her turn comes.
8. Do not make judgments on only one or a few samplings. You may have asked a good student the only question he/she can't answer.
9. Realize that any feedback you have received is limited since all students do not answer all questions.

Questions to Promote Discussion

1. What is the major use which can be made of Oral Testing? Why?
2. What is the chief weakness of Oral Testing? Why?
3. What would you do to compensate for the weakness mentioned in response to Question #2? Why?

PERFORMANCE TESTING

Definition and Description

In Performance Testing, the student is required to perform some physical feat or demonstrate some skill which has been taught. Usually this type of testing is applied to psychomotor learning. Generally students are tested individually although it is possible to measure objectives aimed at team or group skill development.

Advantages or Special Purposes

1. Performance Testing is best used to measure learning objectives concerned with skill development.
2. Learning is directly measured by having students doing rather than writing or talking about doing.
3. As students realize their need to demonstrate what is to be learned, motivation to learn is often increased.
4. Models of performance to be reached often improve the learner's performance.
5. Knowing the performance standard to be reached lets the learner work individually with efficiency to reach the standard.

Disadvantages or Limitations

1. Performance Testing is subjective in scoring. (This is why competitive performances are often scored by several judges, with the high and low scores thrown out.)

2. Lacking objectivity, Performance Testings is thus low in reliability.
3. Performance Testing is low in comprehensiveness because students are generally measured one at a time, so that only a few skills can be measured.
4. It is often necessary to put in a great amount of time to set up a performance test, especially if an apparatus is used.
5. Scoring takes time, especially to translate the performance effectively into a set of meaningful symbols for the learner's benefit.
6. Class time is not efficiently used since the rest of the students are not meaningfully involved while the one student is performing.
7. Discipline problems often arise among those "sideline" students not involved in the performance evaluation being conducted.

Guidelines for Maximum Utilization

1. Use Performance Testing only for learning for which it is valid: skill learning.
2. Demonstrate (or have demonstrated) and explain the standards of acceptable performance.
3. Explain the scoring system to be used and how each factor involved in the performance will be scored.
4. Keep the scoring system as simple as possible and based solely upon observable criteria.
5. Increase the objectivity of the scoring process by training student "verifiers" as scorers.
6. At the time of testing, have all required apparatus pre-arranged and checked out for operability and safety.

7. Define clearly what meaningful activity those students not directly involved in testing should be doing.
8. After the performance, explain to the student the significant strengths and weaknesses of his/her performance.

Questions to Promote Discussion

1. What do you consider to be the best use of Performance Testing? Why?
2. What do you consider to be the major drawback to Performance Testing? Why?
3. What would you do to minimize the effects of the drawback you listed in response to Question #2? Why?

TRUE-FALSE

Definition and Description

Basically, the student is presented with a statement which he or she must decide is either true or false. Teachers use variations of the item in which students must provide the correct answer if the statement is false (true-false corrected).

Advantages or Special Purposes

1. True-False items are best used for those situations where the student is expected to identify the correct alternative from among two possibilities.
2. Scoring this item is highly objective.
3. A great amount of learning material can be sampled in a minimum amount of time, thus contributing to a high degree of comprehensiveness and efficient use of class time.
4. Items are relatively easy to construct, but not as easy as is often believed.
5. Not much time is required for scoring.
6. True-False tests can provide the teacher with a quick means to analyze the degree of general understanding a class has of a large block of material.

Disadvantages or Limitations

1. True-False items are susceptible to guessing. Students have a 50-50 chance of guessing correctly.
2. Because of this susceptibility to guessing, reliability is fairly low.
3. Students seem to be less motivated to study for this type of item, perhaps due to the “guessability” factor.
4. While True-False items are relatively easy to construct, they are often designed without the necessary planning, which results in poorly worded items.

Guidelines for Maximum Utilization

1. Use True-False items only for those situations where two alternatives exist.
2. Use statements that are completely true or completely false. No “trick” questions.
3. Use short, simple sentences.
4. Paraphrase rather than copying directly from textbooks.
5. Avoid arranging an answer pattern which makes it easier to score, but can also be detected by students.
6. Avoid the hints given by statements using “always,” “never,” etc.
7. Aim for approximately one-half true answers and one-half false answers for a test.
8. Use “true-false corrected” and other variations to minimize the effects of guessing.
9. If the statement is an opinion as part of the question, state who is being cited.

Questions to Promote Discussion

Below are questions which contain some errors. Using the above material, identify the errors and explain how the items should have been constructed. (Cite the specific statement in the text whenever possible.)

True-False (Not necessarily parts of a single test)

- 1. Never punt on third down in football.

RANKING OF INHERENT CHARACTERISTICS OF VARIOUS

Testing Means or Items ↓ Criteria	Oral	Performance	
			Essay
Validity	For objectives concerning oral skill development	For objectives requiring student to perform feats or demonstrate skills	For objectives requiring student to develop written skills, interpretation, use data, etc.
Reliability	Low	Low	Low
Objectivity	Low	Low	Low
Comprehensiveness	Low	Low	Low
Time Needed For: Construction / Scoring	Medium / Medium	Low / Low	Medium / Low
Efficient use of classtime	Low	Low	Low
Ease of: Construction / Scoring	Medium / Low	Medium / Low	Medium / Low

- 2. The Big Ten is the toughest football conference.
- 3. Davy Crockett, who died at the Alamo, was once governor of Rhode Island.
- 4. In true-false questions, use short, simple sentences.

TESTING MEANS OR ITEMS, ON A COMPARISON BASIS

Written			
Completion	True-False	Multiple Choice	Matching
For objectives requiring recall. Student is to name, to list, etc.	For objectives which require student to select from 2 alternatives	For wide range of objectives and various levels of thought	For objectives requiring student to match or associate related objects or ideas
Medium	Low	High	Medium
Medium	High	High	High
Medium	High	High	High
High Medium	High High	Low High	Medium High
Medium	High	High	High
High Medium	High High	Low High	Low High