



Three Must-Know Fundamentals of Good Tests

by Jack H. Shrawder, TFS Executive Director

A major portion of your instructional time and effort is focused on preparing, administering, and scoring tests. Three fundamentals are crucial to the construction or selection of good tests or test questions—purpose, validity and length.



Purpose

Testing, like most activities, requires that the purpose be clearly defined from the outset. What is the main reason for testing learners? Probably, you have a good idea, but just to be sure, see if you can answer the following multiple-choice question correctly.

From the following four statements, identify the one that expresses an **inappropriate or wrongheaded** purpose for testing:

- A. To sample what a student now knows and accurately predict how well he or she will perform on questions not asked.
- B. To provide the instructor with feedback on the effectiveness of instruction.
- C. To penalize a class for inattentive behavior during a lecture.
- D. To indicate to students the content that is really important to master.
- E. All of the above.

Since tests can teach as well as measure performance and knowledge, let's analyze one by one, the question (called the stem), the right answer, and the three distractors.

Stem

The stem is a negative statement which adds complexity and makes more likely that some students will misread or not understand what is being asked. In general it is better practice to write positive question stems.

Correct Choice

The most **inappropriate** reason for testing is statement C (Of course, answer C is always the one to choose; right?) Some teachers and students actually do think of tests as punitive devices, but this is an inappropriate purpose, especially when you are working to build a facilitative, rather than an authoritarian, relationship with students.

Distractors

Now, choice A is quite valid and appropriate. Since there is never enough course time to test all knowledge taught, a major purpose of any test is to adequately **sample** a student's knowledge so that any one test score accurately represents what a student knows across a particular knowledge or skill area.

Also, tests provide you with student performance (how well they are learning) data in order to better plan, modify and deliver instruction. Therefore, gathering learning effectiveness data represents a secondary but equally valid purpose for testing—making choice B logical and very appropriate.

Finally, let's consider Choice E, Some test writers like to add "All of the above," or "None of the above," to all of their test questions. At first glance this might seem to be an easy and clever way to increase the difficulty of the test. But in reality, it is poor practice. Including these two

choices can muddy the waters to such an extent that the question no longer measures what the test question writer thought it would. Generally the best approach is to avoid adding these two choices to your exam questions.

Test questions can also serve as pointers to students that highlight what the instructor or course designer believe is the most important material to walk away with. Hence, testing does fulfill the purpose of directing and focusing students' study efforts. Statement D then, also represents a viable and appropriate purpose.

How did you do? Do you agree with the question explanations? Why or why not?

Validity

Next consider, **Validity**. Validity is a term that embodies the concept of **degree of appropriateness**. It's a relative measure of how well the interpretation of a test's score relates to the purpose for giving the test. In other words, a highly valid test will always produce test scores consistent with the purpose for the test.

There are several characteristics that can be examined when you want to assess the validity of your tests. Consider these aspects whether you construct a test by choosing items from a test bank or by building questions from scratch.

Ideally, each test question or set of questions should:

- Measure the breadth and depth of knowledge desired.
- Accurately measure what the question is designed to measure.
- Be constructed in the style, structure, vocabulary, question type and length appropriate to the students being tested.
- Produce a score for each student that closely corroborates with other measures of the student's achievement.

Length

Finally, consider the last major factor, test length. A good question is: How do you know if a test is too long?

A take away concept here is the 90 percent rule. Here's how it works.

A good rule of thumb is to monitor the percentage of students who are able to complete the test before time has expired. Lewis Aiken, author of a comprehensive article on testing, "Testing with Multiple-choice Items," *Journal of Research and Development in Education*, 1987, recommends that at least 90 percent of the students should be able to complete an exam in the time allowed.

The more items you cram into a test the more the test score becomes a reflection of the student's ability to work rapidly and accurately. Choosing fewer items tilts the test toward measuring thinking and work-checking abilities since students have more time to consider each item carefully and can double-check their work and subsequent answer-sheet marking for any errors.

The three fundamental factors to always keep in mind during test preparation are: purpose, validity, and length.