How to Test and Evaluate Learning

Written by
Kay Rooff-Steffen, Coordinator of Humanities, Communication and Fine Arts, TFS
Partner Author, Eastern Iowa Community College District, Muscatine, IA.
Brian R. Shmaefsky, Ph.D., Professor of Biology, TFS Partner Editor, Science and Technology, Lone Star College, Kingwood, TX.
Michael Griffith, PA.
Jack H. Shrawder, Publisher, Teaching For Success, South Lake Tahoe, CA.

Edited by
Michael Griffith, PA.
Jack H. Shrawder, Publisher, Teaching For Success, South Lake Tahoe, CA.

Section 1. Overview

Student arguments, complaints and disappointments over testing are all a possibility during any course, unless you have developed a sound working knowledge of the evaluation and testing skills required to be a successful instructor. Conflict occurs because students care deeply about tests and test results, since they so profoundly affect their academic careers. When I talk with students, I find poorly written, tricky and ambiguous or irrelevant test questions are at the top of the list of students’ complaints about their instructors. Students deserve good, fair tests, and this QuickCourse can help you create them for your students.

This Teaching For Success QuickCourse is designed for you, the busy, caring part- or full-time instructor who desires to develop teaching skills, but must make every minute devoted to professional development count. This self-study course will give you the fundamental knowledge you need to be comfortable and competent in the area of testing and evaluation.

What this course can do for you.

After completing this QuickCourse you should know:
- The purpose of testing.
- Common test and evaluation terms.
- What a test should accomplish.
- How to use a TFS (TPAG) Test Planning and Analysis Grid to check the balance of your tests.
- What question words correspond to various levels of learning achievement as described in Bloom’s taxonomy.
- How to evaluate test questions.
- How to estimate the proper length for a test.
- How to create a Course Testing Plan.
- How to use the TFS Course Testing Planning Grid to create a good course testing plan.
- Resources for more detailed testing information.
Is this Quick Study for me?

The first principle of effective test design is overcoming the notion that testing merely involves writing a set of questions. In truth, it starts with specific, defined course outcomes that are linked to effective lesson plans and based on a sound learning model such as Bloom’s Taxonomy.

Good testing is an outgrowth of a well-organized course that helps students study, understand, apply and retain what has been learned. Therefore, tests should never be constructed as a guessing game for students.

Applicability

This Quick Course is designed for the instructor in any discipline who is faced for the first time with the tasks of planning, creating or selecting test questions and test types and administering tests in their classrooms.

Time required

You should devote about 60-90 minutes to reading and interacting with this QuickCourse. You will be asked to answer questions, then practice applying this new knowledge to improving a test you created or selected from a test-question bank. You can practice writing test questions in interactive text fields. If time is short, just take it one section at a time.

Instructional design

The learning process in this QC starts with a consideration of the three functions of tests, then moves on to creating a testing plan based on the six learning levels described by Bloom’s Taxonomy.

Rules for test writing are reinforced with sample objective-style questions demonstrating good and bad question construction techniques. Finally, test-length factors are considered, a course testing plan is presented and links to several testing websites are provided.

Main Study Topics

- Testing talk: What do common testing and evaluation terms mean?
- Planning a test: What do you want the test to do?
- Question construction: How do you write or select good, clear, concise and valid test questions?
- Test length: How to estimate the length of a test.
- Course testing plan: How to create a course testing guide.
- Self-test: Check your understanding of the terms, recommendations and principles contained in this QuickCourse.
- References: Book and articles that can provide more testing details.
- Job aides: Blank TFS test design and planning forms.
Section 2. Testing Basics

Since childhood, for many of us, the word “test” or “exam” has been a stress-producing trigger word. Tests are dreaded by some students and faculty alike. Therefore, testing is an emotionally-charged subject for students.

Poor tests and questionable questions will evoke more student disagreements and arguments than any other teaching and learning activity. This QuickCourse will help you improve your tests and your students’ satisfaction with the way you evaluate them.

You may assume that since you have taken a plethora of tests throughout your academic career that you know how to create a good, fair test. If so, you have discovered or are about to discover that testing and evaluation is a complex subject and demands learning a body of knowledge and applying some practical skills.

You have experienced (or if new to teaching, about to experience) the grief that poorly constructed or planned tests can cause students and instructor alike. Without a basic familiarity with good testing principles, you may feel the burden of wondering whether or not your tests adequately evaluate student knowledge and comprehension. This QuickCourse will give you the tools so that you can know for sure that you are doing a professional job in testing and evaluation.

Tests should never be seen by your students as a guessing game or an exercise in interpreting ambiguous questions. Questions should be clear, straightforward and understandable. Good tests begin with good learning objectives.

Learning objectives

One of the most logical ways to look at testing is to recognize that test items are really learning objective statements transformed into questions.

Quality tests depend on quality learning objectives. If you don’t know or can’t define or express precisely what students must know, do or choose, how can you develop accurate tests to measure learning achievement? So job-one is to be able to write or identify learning objectives. If you are new to teaching and have not been given a list of learning objectives or expected outcomes for your course, ask your department head if such a list exists. If you fail to find a prepared list for your course, check the course textbook. Some instructor’s text editions will list learning objectives.

A formal learning objective statement has three parts:

- Starting conditions. (Where do you start?)
- Action. (What will the learner do with the learned knowledge, skills or attitudes.)
- Performance standard. (How well will the learner do it?) Possible time or accuracy standards stated.
Example three-part Learning Outcome

Given (starting conditions) a PC with a text editor without a spell checker installed, the student will be able to (what learner will do) write a paragraph in five minutes consisting of a topic sentence, and four supporting sentences.

The paragraph will (performance standard—how well) conform to standard grammar rules with no more than two grammar or punctuation errors and have no more than one word misspelled.

Granted that writing a formal, three-part learning objective (Given, Do, Standard) at this level of detail takes considerable time and effort and therefore, may be an impractical task for many part-time faculty. But the principle of specifying learning outcomes is a sound instructional design and testing principle. To ensure quality testing, it is essential that you have a list of learning objectives or learning outcomes available prior to the preparation of each test.

Not an option

For most of us, we are forced to test by our institutions and federal or state licensing boards or agencies. Paper or computer objective and subjective tests are the most common form of evaluation used today at all educational levels. So, whether to test is not usually an option, but you do have the option to decide how and oftentimes when to test and what the results mean.

Some instructors rely solely on tests extracted from an instructor’s edition of a textbook, but these pre-made tests are geared toward the average student taking an average version of the course. Therefore, they may not ensure quality testing unless the questions are evaluated and selected according to a criteria based on a specific course and student group.
Since you are an individual and unique teacher, chances are your classes reflect your specific strengths and experiences. Your tests should therefore reflect your special approach, and often, the only way to accomplish this is to create tests from your own test questions.

Whether you are faced with the task of writing test questions from scratch, selecting questions for a test from a bank of publisher-supplied test questions or choosing a test from those offered by your textbook publisher, the information in this QuickCourse will help you evaluate your students with a high quality test.

**Your turn**

Describe in a few brief sentences in the text field below the most pressing problems that you face with testing and evaluation.

Why is a list of learning objectives or learning outcomes essential to have before a test is created? Write your answer in the field below.

Name the three parts of a formal learning objective statement?

- Part 1.
- Part 2.
- Part 3.

Answers?

Click here to go to answers.
Section 3. Testing Talk

How’s Your Testing Talk?

Testing theory and practice has its own special terms that may need some getting used to. If you are new to testing and test terminology, you’ll find it helpful to be comfortable with the terms when conversing with fellow faculty, administrators or for easy reading of reference materials. These are testing and evaluation terms you should know: (term links are active):

- Evaluation
- Formative evaluation
- Summative evaluation
- Diagnostic evaluation
- Test
- Measurement
- Performance
- Standard
- Reliability
- Validity
- Norm-referenced
- Criterion-referenced
- Objective-referenced
- Objective question
- Subjective question
- Sample

Here are the simplified, working definitions that we use at Teaching For Success: Use these terms appropriately and you will gain respect as a competent instructor!

Evaluation—This is the granddaddy term that means using a systematic process to make a sound judgments about the value of something.

Formative evaluation—This type of evaluation is used to create a progress report during a course of learning. Quizzes, midterm exams, chapter take-home tests, journal reviews and portfolio critiques could all be used in a formative evaluation.

Summative evaluation—One subdivision deserves another. An evaluation that sums up the outcomes of a learning process at the completion of the learning is a summative evaluation. A final exam is usually part of a summative evaluation.

Diagnostic evaluation—And, there is even another subdivision. This term describes a type of evaluation that is all too often neglected in the higher education classroom. It is used to pinpoint learning problems or uncover gaps in knowledge or skills that are preventing the student from progressing.
Test—In testing circles, tests are referred to as **instruments** that measure how well a learner performs.

**Measurement**—When you measure something, you get the results in numbers and units for example, a piece of string with a length of 9 inches. Similarly, when you **measure** learning you use an instrument that provides a numerical result.

**Performance**—Now, this is a term that really piques the interest of an Educational Behaviorist. Performance is about results: How well? How soon? How much better? These are good performance questions. To be meaningful, performance should be clearly defined and measured against a standard.

**Standard**—A most important testing concept. When you think about standards in testing you are pondering the question, **How well** must my students perform and what will be the measuring stick? You can create a test, but without setting or choosing a standard, the results are meaningless. Standards, for example, may be expressed in percentage correct, or as a number or items or actions to be completed in a specified amount of time or how accurately work is accomplished compared to a fixed model or list of tasks, or compared to others performing the same action in the same class or in all course sections.

**Reliability**—This term describes how well a test would produce the same results if given to other groups of students at different times and places. A reliable test would yield approximately the same range of scores and share similar mean scores whenever it’s given. When you create your own tests you can expect that their reliability will be low. The same test may yield varying scores from use to use.

**Validity**—This term is crucial to understand. Valid tests measure what they are designed to measure.

*Valid*—This test is valid. It measures what it was designed to measure.
The importance of validity

Unless your tests are carefully planned, designed and constructed, they will have low validity and fail to measure the knowledge, skills or attitudes that you intend. If your tests lack validity, expect arguments and challenges over correct responses and what the question is really asking. When invalid tests are used to make important grading and competency decisions, tragic errors can result. Valid tests are good, fair and accurate tests of the knowledge and skills to be assessed.

How to Maximize Validity

- Ask a peer, friend or spouse to review the questions and identify ones that are confusing or ambiguous.
- Check question construction according to the construction principles found in the “Question Construction” section of this course.
- Use a test construction grid to match questions with learning objectives. Eliminate any superfluous questions.
- Ensure students know precisely how to answer each question as well as the scoring and grading criteria.

Norm-referenced—A norm-referenced test measures the performance standing of an individual in reference to the performance of a group. When you grade on the curve you are norm-referencing the test.

Criterion- or objective-referenced—This type of test measures the performance of a student against a defined set of learning tasks or list of learning objectives. Mastery learning uses criterion-referenced tests to evaluate whether a student has mastered an acceptable number of learning objectives.

Objective question—Objective questions are those that can be scored without a detailed analysis of the answer. Multiple-choice, fill-in-the-blank, true-false, matching and one-word short answer are all examples of objective test questions.

Subjective question—Objective questions are those that must be scored by detailed analysis and repetitive scoring process; whereas, essay test questions are subjective questions. The answers to subjective questions are subject to the examiner’s opinion on the correctness of the response.