

How do I Create Tests for my Students?

Mekiva Callahan & Micah Meixner

Introduction

Assessments are essential to the learning process. As the class instructor, you need a means of gathering information on the effectiveness of your instruction and a way to measure your students' mastery of the course's educational outcomes. For students, assessment provides them with feedback on their learning and can also be an incentive for improving academic performance. From an administrative perspective, the cumulative value of assessments is tangible data suggestive of student achievement. Perhaps the most well known form of assessment is a test or an exam, so given the high stakes of evaluation, from a variety of perspectives, and the importance of accurately gauging students' learning, it's imperative that you design valid, well-written tests.

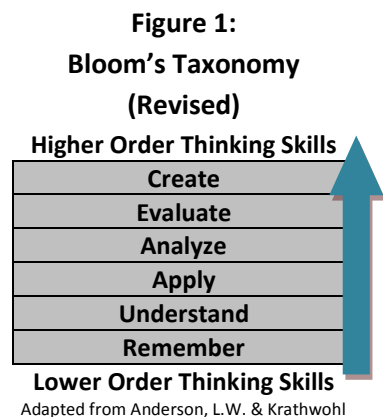
Test Composition and Design: Where to Start?

Identify Course Goals

It might seem obvious, but one of the most important steps of test composition is to revisit your overall goals and objectives for the course and to determine which goals you intend to evaluate with this test, bearing in mind that a formal test or exam is not always the best way to evaluate the desired learning outcomes. Once you have identified which outcomes what you want to measure, consider what type of question or prompt best facilitates the students' production of that outcome. For instance, is it your intention to create a test asking students to recall definitions or are you interested in having students demonstrate their ability to compare various concepts and defend their position on a controversial subject?

Many people utilize Bloom's Taxonomy, a hierarchical structure of thinking skills, as a tool for gauging the cognitive depth of student learning. Figure 1 depicts a more recent adaptation of Bloom's Taxonomy and it can be useful to keep in mind when constructing tests. Consider, for instance, the difference between the first test or assignments of the semester when you simply want to measure the students' ability to understand or recall new information, and the final exam in which you ask students to independently analyze data or situations or possibly create a project or document representative of the information covered throughout the semester. The required skills or desired level of cognition will vary based on the educational objectives for each exam, so it is vital that you keep your pre-determined goals and objectives in mind throughout the test composition progress.

For additional information on Bloom's Taxonomy and sample questions for each level of cognition, please see the additional resources provided.



Determine Test Structure/Design

Much like learning styles, research shows that many students have a preferred test format, so in order to appeal to as many students as possible you might consider drawing from a variety of testing methods or styles. In fact, you can design a single exam to include several kinds of questions and measure a range of cognitive skills. Some common types of tests and test items are discussed below.

- **Objective Tests:** An objective test is one in which a student's performance is measured against a standard and specific set of answers (i.e. for each question there is a right or wrong answer). When composing test questions, it is important to be direct and use language that is straightforward and familiar to the students. In addition, the answer choices provided on the test should be challenging enough that students aren't able to guess the correct answer simply by comparing how all of the options are written. Examples of objective test items include the following.
 - Multiple-choice
 - True-false
 - Matching
 - Problem based questions. These require the student to complete or solve an equation or prompt and are commonly used in application based courses such as mathematics, chemistry, and physics.

For further information on writing effective test questions, please contact the TLTC for a consultation or refer to the online resources provided below.

- **Subjective Test:** Unlike objective tests for which there is a definitive standardized or formulated answer, subjective tests are evaluated based on the judgment or opinion of the examiner. Tests of this nature are often designed in a manner in which the student is presented with a number of questions or writing prompts for which he/she will demonstrate mastery of the learning objective in his/her response to the question. When composing prompts as test questions, it is crucial that you phrase the prompt clearly and precisely. You want to make sure that prompt elicits the type of thinking skill that you want to measure and that the student's task is clear. For example, if you want students to compare two items, you need to provide or list the criteria to be used as the basis for comparison. Examples of subjective test items include the following.
 - Essay
 - Short answer

When grading subjective tests or test items, the use of an established set of scoring criteria or a well-developed rubric helps to level the playing field and increase the test's reliability. For more information on rubric development, please see the additional online resources provided.

Table II contains a chart showing advantages and disadvantages for a selection of test items. It's important to note that this is not an exhaustive list, and remember that as the course instructor, you have the freedom to choose what form of assessment most aptly measures your specific learning objective.

Table II: Advantages and Disadvantages of Commonly Used Types of Achievement Test Items

Type of Item	Advantages	Disadvantages
<i>True-False</i>	Many items can be administered in a relatively short time. Moderately easy to write and easily scored.	Limited primarily to testing knowledge of information. Easy to guess correctly on many items, even if material has not been mastered.
<i>Multiple Choice</i>	Can be used to assess a broad range of content in a brief period. Skillfully written items can be measure higher order cognitive skills. Can be scored quickly.	Difficult and time consuming to write good items. Possible to assess higher order cognitive skills, but most items assess only knowledge. Some correct answers can be guesses.
<i>Matching</i>	Items can be written quickly. A broad range of content can be assessed. Scoring can be done efficiently.	Higher order cognitive skills difficult to assess.
<i>Short Answer or Completion</i>	Many can be administered in a brief amount of time. Relatively efficient to score. Moderately easy to write items.	Difficult to identify defensible criteria for correct answers. Limited to questions that can be answered or completed in a few words.
<i>Essay</i>	Can be used to measure higher order cognitive skills. Easy to write questions. Difficult for respondent to get correct answer by guessing.	Time consuming to administer and score. Difficult to identify reliable criteria for scoring. Only a limited range of content can be sampled during any one testing period.

SOURCE: Table 10.1 of Worthen, et al., 1993, p. 261.

Test Composition and Design: Additional Considerations

Validity & Reliability

Two key characteristics of any form of assessment are validity and reliability. As Atherton (2010) states, “a valid form of assessment is one which measures what it is supposed to measure,” whereas reliable assessments are those which “will produce the same results on re-test, and will produce similar results with a similar cohort of students, so it is consistent in its methods and criteria.” These attributes provide students with the assurance they need to know that the test they are being given is fair and reflective of what has been covered in the course.

To establish a valid test instrument it is important always be mindful of your pre-determined learning outcomes and goals. This mindfulness will help to ensure that each question you develop is an accurate measure of the specified learning outcome. An example of an invalid question is one which tests a student’s ability to recall facts when it was actually intended to assess a student’s ability to analyze information.

As Atherton (2010) describes it, another way to think of reliability is in terms of “replicability.” Is there a general consistency in students’ overall performance on an exam? If the exam is given to more than one class or over the course of multiple semesters, is there consistency between the various classes? If so, the test is considered to be reliable. Strategies such as writing detailed test questions or prompts, including clear directions, and establishing and communicating clear grading criteria will increase test reliability.

Verifying validity and reliability in a written test can be challenging. For instance, in the grading of a writing test, what exactly are you trying to measure? The students’ writing abilities, content knowledge, all of the above? Be sure that all of your expectations for an exam are communicated to your students well in advance and always be sure that your expectations mirror those articulated in the overall course goals. Another way to increase test validity and reliability is to reexamine and possibly remove questions missed by a large majority of students. If a significant percentage misses the same question, there is a definite possibility that the question was somehow unclear or was not representative of the intended learning outcome.

Test Length:

Another important aspect of test composition is time management—on the part of the professor as well as the student. A common student complaint with tests is that the test was covered material never covered in class or had too many questions on something that was covered in only a few minutes. When designing tests it is helpful to remember that topics on which you spent a significant amount of class time, through instruction and activities, should be appropriately emphasized on the test. This does not mean that you should not include items that received less coverage in class, just be sure to maintain an appropriate balance.

Also, bear in mind that it will take students longer to complete the test than it would you. In his highly referenced book *Teaching Tips* (1994), Bill McKeachie outlines the following as a strategy for determining test length, “I allow about a minute per item for multiple-choice or fill-in-the-blank items, two minutes per short-answer question requiring more than a sentence answer, ten or fifteen minutes for a limited essay question, and a half-hour to an hour for a broader question requiring more than a page or two to answer.”

Conclusion

Tests and exams often play a significant role in the overall assessment of students’ learning. Therefore, as instructors, it is essential that we pay particular attention to the manner in which we construct these instruments. Remember to always keep your course goals and learning objectives at the forefront of your mind as you begin to determine what kind of test is the best measure of your students’ learning. To that end, if it fits with your course design and content, you may want to consider alternate forms of assessment such as group projects, student portfolios or other activities that extend and build throughout the course of the semester. These alternative or non-traditional forms of assessment frequently offer students a more authentic opportunity to apply their knowledge and higher-order thinking skills.

Creating tests and other forms of assessment can be a challenging task, but there are plenty of resources available to you. If you would like assistance with test composition, or if you have questions about assessment in general, please contact the TLTC for a consultation.

Online Resources

Bloom's Taxonomy:

Rathburn, S. Teaching Tip Bloom's Taxonomy: Testing beyond Rote-Memory. Colorado State University, The Institute for Teaching and Learning

<http://tilt.colostate.edu/tips/tip.cfm?tipid=87>

This website provides a general and succinct overview of Bloom's Taxonomy with sample test questions from each of the cognitive levels established in Bloom's Taxonomy.

Vanderbilt Center for Teaching

http://www.vanderbilt.edu/cft/resources/teaching_resources/theory/blooms.htm

This link to Vanderbilt's Center for Teaching contains detailed information on Bloom's taxonomy, both the original and the revised taxonomy. It defines each level and provides a list of useful verbs to describe the cognitive processes associated with each level.

Writing Test Questions:

"Simple Guidelines for Writing Test Questions." Renton Technical College Office of Instructional Improvement

<http://webs.rtc.edu/ii/Teaching%20Resources/GuidelinesforWritingTest.htm>

This website offers a comprehensive overview of how to construct test questions, covering a range of questions including True-False, Multiple-Choice, Matching, Completion, and Essay questions.

<http://web.uct.ac.za/projects/cbe/mcqman/mcqappc.html#C1>

This is the appendix from the web version of Designing and Managing Multiple Choice Exam Questions, a handbook for the University of Cape Town, South Africa. This link provides examples on how to write questions that are on the higher level of Bloom's as well as more application of each level.

Clegg, V.L. & Cashin, W.E. (1986). "Improving Multiple-Choice Tests." Idea Paper 16 from Kansas State University's Center for Faculty Evaluation and Development.

http://www.theideacenter.org/sites/default/files/Idea_Paper_16.pdf

Cashin, W.E. (1987). "Improving Essay Tests." Idea Paper 17 from Kansas State University's Center for Faculty Evaluation and Development.

http://www.theideacenter.org/sites/default/files/Idea_Paper_17.pdf

For detailed information specific to improving multiple-choice and essay tests, take a look at these two IDEA papers. These are links to pdf files, and each contains information on constructing test items, the strengths and limitations for each item, and recommendations on using each to assess student learning.

Assessment: Rubrics, Validity & Reliability

Atherton, J.S. (2010). Learning and Teaching; Assessment [On-line] UK

<http://www.learningandteaching.info/teaching/assessment.htm>

In this article, Atherton provides an accessible discussion of assessment, applicable to any discipline or type of course.

Rubistar.

<http://rubistar.4teachers.org/>

This website offers a free tool to help instructors create their own rubrics. It also offers a search tool to find sample rubrics.

Davis, B.G. (1993). "Quizzes, Tests, and Exams" and " from Tools for Teaching. Accessed online <http://teaching.berkeley.edu/bgd/quizzes.html>

A great reference for faculty, Barbara Gross Davis's book Tools for Teaching, has several chapters devoted to the topic testing and grading. This is a link to the chapter on assessment. It offers a broad overview the topic as well as some general strategies for test design. The hard copy, which is available in the TLTC library, has specific chapters on designing all the items highlighted in this article.

<http://www.park.edu/cetl/quicktips/authassess.html>

Park University maintains an excellent website on summative assessments. This particular link compares authentic assessment to the traditional form as well as lists advantages and disadvantages of using this to assess student learning. In addition, it provides ideas for implementing various types of authentic assessments into your course.

Additional References

Anderson, L.W., and Krathwohl, D.R. (2001). A Taxonomy of Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives. New York: Longman.

Bloom, B.S., et. al. (1956). Taxonomy of Educational Objectives: Cognitive Domain. New York: David McKay.

Gronlund, N.E., & Linn, R.L. (1990). Measurement and Evaluation in Teaching (6th ed.). New York: Macmillan Publishing Company.

Improving Multiple Choice Questions. (1990, November). For Your Consideration, No. 8. Chapel Hill, NC: Center for Teaching and Learning, University of North Carolina at Chapel Hill.

Mandernach, B. J. (2003). Multiple Choice. Retrieved June 21, 2010, from Park University Faculty Development Quick Tips.

McKeachie, W. J. (1994) Teaching tips: Strategies, Research, and Theory for College and University Teachers (9th ed.). Lexington: D.C. Health and Company.

Svinicki, M.D. (1999a). Evaluating and Grading Students. In Teachers and Students: A Sourcebook for UT Faculty (pp. 1-14). Austin, TX: Center for Teaching Effectiveness, University of Texas at Austin.

Worthen, B. R., Borg, W.R., & White, K. R. (1993). *Measurement and evaluation in the schools*. New York: Longman.