

ASSESSING STUDENT LEARNING

RANDOLPH A. SMITH
LAMAR UNIVERSITY



BOOKLET

13

عمادة تطوير المهارات
إنجاز متميز .. والتزام بالتطوير





Assessing Student Learning

Randolph A. Smith

Lamar University



عمادة تطوير المهارات
Deanship of Skills Development

جميع حقوق الطبع محفوظة
عمادة تطوير المهارات
١٤٣٤هـ - ٢٠١٣م

Assessing Student Learning



Although many faculty truly enjoy teaching, one of the necessary evils of teaching is assessing student learning and assigning grades. I have often heard faculty remark that teaching is great except for the grading aspect. Based on the number of student complaints about grades that faculty receive, it is doubtful that students are any more positive about the grading aspect of school than are faculty. Nevertheless, it is an imperative for faculty to assess student learning and give grades on the basis of those assessments. In this booklet, I focus on the process of assessing student learning and not on grading-I believe that the better job we as faculty do at assessing student learning, the less onerous the grading process becomes.

There are several important decisions faculty members must make as they plan assessments of their students' learning. Although there is no prescribed order in which faculty must make these decisions, I will cover them in the order that seems like a logical starting point to end point.

Formative Versus Summative Assessment

Summative assessment refers to what I described in the opening paragraph-assessing student learning for grading purposes. This type of assessment is the process with which students and faculty are most familiar. It is also the type of assessment that tends to have negative emotions associated with it, again involving both students and faculty. Typically, summative assessments take place

Assessing Student Learning

at the end of a learning process—a class unit, a project, an entire course. Because of this temporal arrangement, students may not benefit much in terms of their learning from summative assessments. The ultimate example of not benefitting from a summative assessment is a final exam in a course. At most, students may find out their grade on the exam but rarely get any feedback other than that grade. Depending on when a project such as a paper is due and returned (if at all), students may also not get any feedback that will benefit their learning. Although students can probably expect to get some feedback from exams that they take early in a semester, the amount of feedback can vary tremendously, as does the degree of attention that students pay to that feedback. Thus, one of the major problems with summative assessment is that the feedback process may not benefit student learning very much, if at all.

On the other hand, formative assessment is designed solely for the purpose of providing students with feedback. For example, an instructor might give students daily quizzes over the material but not use performance on the quizzes as part of the grade for the course. In this manner, students are getting formative feedback about how well they know the material as they go through the course, but the assessment is not summative because it does not contribute toward their grade. Students can use information from formative assessments to get feedback about how well their approach toward the course (e.g., reading, studying, processing the material) is going. If the formative feedback is not good, students can alter their approach to the course before they take a summative assessment that counts toward the grade. By the same token, faculty can use the results from formative assessments to alter the way they are approaching the course (Cangelosi, 2000). For example, if I am teaching a course and use a formative assessment process, I might find out that all the students perform very poorly on a specific topic. Based on this feedback, I might decide that I want to alter my teaching approach in hopes of helping students learn more or better. In contrast, I might decide that my assessment instrument was too difficult or did not match well with the manner in which I was teaching. Finally, I might determine that students did not take the formative assessment measurement very seriously and that I do not need to change anything about how I am teaching. This last conclusion might be appropriate if some of the students performed well on the formative measure and others did not. That information would tell me that the assessment measure I used was not unreasonably difficult given that some students performed well on it. I might want to conduct further diagnosis with the students who performed poorly to determine if I can learn why they performed so poorly.

Another advantage of formative assessment is that, unlike summative assessment, the instructor does not have to spend a great deal of time in grading. Multiple-choice items can diagnose whether students have learned the factual information that is often vital to their later success in the course. If instructors use a written formative assessment, they can use the information gleaned from those measurements to alert the class to common misconceptions or misunderstandings that show up in the assessment. For example, if students are supposed to use a specific style or format for their writing (e.g., American Psychological Association, Modern Language Association, Turabian), a formative assessment assignment may show that many students are making the same mistakes. Rather than marking every student's paper and returning each one individually, the instructor could compile a list of common formatting mistakes and give a copy to each student, thus saving a great deal of grading time.

A final advantage of formative assessment is that instructors may use this approach more frequently in a course than summative assessment. For example, I have used student response systems (also known as clickers) in my classes for several years. One simple method of formative assessment using clickers is to begin each class with a few review questions from the previous class (or classes). Students can quickly record their answers with the clickers, and I can then display the results to the entire class. Students can quickly determine how well they learned the material from the previous class. In addition, I can use the feedback to determine whether there are any widespread misconceptions among the students—if there are, I can go over the misunderstood topic again before moving on. If you do not have access to clickers for your classroom, you can provide students with the same benefit by projecting questions via PowerPoint or an overhead projector, or you can even print them on slips of paper (however, unless you ask for a show of hands, you will lose the benefit of being able to see the entire class's results immediately.). Given that formative results are primarily for the benefit of the students, you can allow students to score their own questions.

In light of these two approaches to assessment, one of the first decisions a faculty member needs to make is which approach to use. Clearly, some summative assessment must take place in order to give grades. Also, there is no requirement to use any formative assessment whatsoever. However, if a faculty member's primary goal is improving student learning, some amount of formative assessment is a good idea.

Assessing Student Learning

High-Stakes Versus Low-Stakes Assessment

High-stakes versus low-stakes assessment is quite similar to summative versus formative assessment. In high-stakes assessments, ultimate student success is solely or largely dependent on the test. Thus, a high-stakes assessment would occur if a student has to make an acceptable score on an exit exam in order to graduate from college. With this example, the link between high-stakes and summative assessment should be clear. In comparison, a low-stakes assessment is one that plays no role or only a small role in students reaching a goal. Thus, all formative assessments fall into the low-stakes category; however, a summative assessment such as a quiz that makes up only a small portion of a student's grade can also be a low-stakes assessment.

Faculty members must make determinations about the relative weight of the assessments they use in classes. The fewer summative assessments used, the more high stakes they become. The larger the number of summative assessments used, the more they move toward lower stakes. However, it is difficult to conceive of an exam or a term paper as ever being considered low stakes by the students who have to complete the assignment. Related to this issue, research studies show both higher student preference for and better performance with more frequent testing (Abbott & Falstrom, 1977; Bangert-Drowns, Kulik, & Kulik, 1991; Peckham & Roe, 1977). These results seem to show that students prefer low-stakes rather than high-stakes assessment so that each assessment has less weight in determining the final grade.

Considering Student Learning Outcomes

After you have determined the type of assessment you wish to conduct, you are almost ready to begin working on the assessment instrument itself. First, however, you must decide *what* you want to assess—what do you want your students to have learned? Ideally, you have already developed student learning outcomes for your course or the relevant unit of the course. If you have not developed learning outcomes, you should read Smith's (2011) booklet on this topic.

Three key points from Smith's (2011) booklet are worth a reminder at this point. First, your learning outcomes must follow from your learning goals. Remember that learning goals can be broad and general, but learning outcomes must be specific. Second, your student learning outcomes must be stated in measurable terms. You cannot develop an assessment of student learning on something that you cannot

measure. Third, learning outcomes are typically linked to a level or type of student learning. Bloom's (1956) taxonomy is the most common example of different levels of student learning, but there is also an updated version (Anderson & Krathwohl, 2001) as well as a digital version of Bloom's taxonomy (Churches, 2008). Remember that the level at which you have students learn information determines the level at which you can assess that information. If you have asked students only to memorize a list of famous people in your discipline, then you should not expect the students to be able to analyze the work of one of the people on the list or to compare and contrast the work of two people on the list. It is important to remember that the different levels of learning are represented by different verbs—verbs that should appear in your learning objectives. Thus, you must assess in a manner that is consistent with the verbs you used in your learning objectives.

There are many different types of assessments; one type of classification scheme that divides them into two groups is objective versus nonobjective assessments (Suskie, 2009). Although Suskie (2009) used the term “subjective” rather than nonobjective, I prefer to avoid that term because it implies possible favoritism or bias in scoring. An objective assessment is essentially any test that can be computer scored, such as multiple-choice, true-false, or matching questions, plus fill-in-the-blank questions. Objective assessments have right or wrong answers such that each question is scored on an all-or-none basis; partial credit for objective items is rare or nonexistent. In contrast, nonobjective assessments involve tasks such as writing, research, or some other task completion. However, as Suskie noted, it is not correct to associate objective assessments as being quantitative and nonobjective assessments as being qualitative. Many nonobjective assessments yield quantitative data.

It is important that you match the type of assessment to your learning objectives. Objective assessments tend to be most useful for measuring student performance at the lower levels of Bloom's (1956) or other taxonomies. Therefore, if you are introducing your students to large amounts of material they have never before encountered, objective assessments such as multiple-choice or true-false tests allow you to quickly and easily assess how much material they have learned.

However, it is exceedingly difficult to write, for example, multiple-choice questions to assess student behaviors such as analyzing, evaluating, or creating. Thus, if you wish to assess these higher order thinking skills, you will likely have to use nonobjective assessments. There are many types of nonobjective assessments to choose from; you can have your students complete items such as writing assignments, essay tests,

Assessing Student Learning

book or research reports, or projects that require them to use material they have previously learned in new and unique combinations, just to name a few possibilities.

To objective and nonobjective assessments, I wish to add a third category that can actually include both objective and nonobjective approaches: learner-centered assessments. Learner-centered assessments are also known as classroom assessment techniques (CATs; Angelo & Cross, 1993). I prefer the term learner-centered, however, because not all assessments that take place in the classroom are aimed at the learner. Angelo and Cross (1993) described their version of classroom assessment as learner-centered because it is aimed at modifying behaviors of the learners rather than the instructor. Changing student behaviors is more efficient than attempting to change faculty behaviors because students who learn new ways to learn can use them in any class, regardless of what the instructor does. One major target of learner-centered assessments is the attempt to modify students' metacognitive skills—"skills in thinking about their own thinking and learning" (Angelo & Cross, 1993, p. 4). This discussion of learner-centered assessments may remind you of the topic of formative assessment discussed earlier in this chapter. Learner-centered assessments are, indeed, aimed at the formative process. Angelo and Cross specifically labeled their CATs as formative in nature. If you are interested in using learner-centered assessments in your classroom, Angelo and Cross provided 50 different ideas for doing so in their book; Table 1 provides examples of their assessments. One nice feature of the book is that Angelo and Cross provide the level of their assessments categorized according to Bloom's (1956) taxonomy.

Table 1

Examples of Learner-Centered Assessments from Angelo and Cross (1993)

Minute Paper

A few minutes before class ends, stop and ask students to write answers to two questions: "What was the most important thing you learned during this class" "What important question remains unanswered?" (Angelo & Cross, 1993, p. 148)

Not only does the first question assess recall, but it also forces students to evaluate the information they received in class. The recall aspect, of course, allows the instructor to determine whether students are understanding the information correctly or if they have misconceptions. The second question provides the instructor a good place to begin teaching at the next class meeting.

Muddiest Point

Ask students “What was the muddiest point in -?” (Angelo & Cross, 1993, p. 154). You can fill in the blank with a variety of stimuli: class, lecture, chapter in text, film, assignment, and so on. This learner-centered assessment provides the instructor with feedback about what students have found or are finding difficult to learn. Armed with this information, instructors have a much better idea of what information to emphasize in class rather than by guessing.

Directed Paraphrasing

The instructor asks students to put some important concept into their own words, usually directed at a specific audience or for a specific purpose and typically avoiding the professional jargon of the academic discipline. This learner-centered assessment makes students go beyond simple memorization and regurgitating information on an exam. If students do not truly understand the material, they will have a difficult time rephrasing it.

Example from a Database Systems (Computer Science) course:

In plain language and in less than five minutes, paraphrase what you have read about computer viruses-such as the Michelangelo virus-for a vice president of a large insurance firm who is ultimately responsible for database security. Your aim is to convince her to spend time and money “revaccinating” thousands of workstations. (Angelo & Cross, 1993, p. 233)

Designing Assessment Instruments

After you have chosen the student learning outcome you wish to assess and have decided the level at which you want to assess it, you can actually craft your assessment instrument. Although this step is certainly the most important one for developing your assessment measure, it should be almost anticlimactic at this point-*if* you have followed all the previous steps faithfully. The reason that actually devising assessment items is anticlimactic is that all you need to do at this point is write items that follow the specific learning objective you have chosen, taking into account the verb you used and the level of the learning that you expect from one of the learning taxonomies, both so that you write the correct type of item.

Suskie (2009, p. 167) recommended beginning to write test items based on a “test

Assessing Student Learning

blueprint”. A test blueprint is simply an outline for your test that includes all the learning objectives that you want students to know for the test. Using a test blueprint ensures that you will not forget to include an important objective that you meant to cover on the test. Likewise, using a test blueprint would let you know if you were writing test items over material that you did not consider as important as your actual learning objectives. Anyone who has compiled many tests during a teaching career without using a test blueprint knows how easy it is to make one of these mistakes. Using a test blueprint will help you allocate items on the test in terms of their importance to your learning objectives rather than being tied to the textbook. For example, perhaps you consider the opening chapter of your textbook to contain less important information than Chapters 2 and 3, which you are also including on your first exam. Based on your learning objectives and a test blueprint, you will include more exam questions from Chapters 2 and 3 than from Chapter 1. Without such guidance, however, you might develop a test that has 10 items from each chapter—in other words, a test that does not match your learning objectives very well.

Finally, using a test blueprint can help you avoid a student criticism that some exams evoke: “This test doesn’t seem to cover what we learned in class”. As long as you have communicated your learning objectives to your students, have taught information related to your learning objectives, and have followed your learning objectives in constructing the exam, there should not be a mismatch.

Writing Good Multiple-Choice Items

Suskie (2009, pp. 170-171) and Hales and Marshall (2004, pp. 65-88) presented a list of tips for writing good multiple-choice items compiled from testing experts and many studies. Bear in mind, however, Suskie pointed out that it is difficult to follow all guidelines without ever violating some of them. Thus, you should look at these ideas to see how your multiple-choice questions stack up. If you find yourself to be a frequent violator of any of these recommendations, it would be wise to try to reduce your violations.

- Write your items as briefly as possible—longer items with excess information are more likely to confuse students.
- Use vocabulary that is as simple as possible (unless you are testing for vocabulary)—all students should have an equal chance on items regardless of language skill.
- Avoid using questions that are interrelated—a student should not be able to use information from one question to answer another question, nor should missing one item cause a student to automatically miss another item.

Assessing Student Learning

- Make the item stem a complete question-students should not have to read the answer options to understand the question being asked.
- Avoid items that ask “which of the following”-slower students may be penalized.
- Avoid items that rely on common knowledge-students should have learned the material in your class, not from previous experience.
- Try to avoid negatives in the stem; if you must use such items, emphasize the negative word (e.g., NOT, **not**, not)-anxious students may read over the negative word and not see it.
- Make sure all options are grammatically correct-avoid giving clues to correct answers through grammar.
- It is not necessary to have the same number of options for every question-if there are only three plausible options (e.g., “goes up,” “goes down,” “stays the same”), then use only those.
- Arrange the answer options logically-if there is a logical order to the options (e.g., numbers that increase or decrease, arrange words alphabetically), use it to make it easier for students to locate the correct answer if they know it.
- Keep the answer options approximately equal in length-students who are good at taking tests know that a longer answer is often correct.
- Avoid the “none of the above” option-a student may know incorrect answers but not the correct answer; if you do use it, use it more than once as both incorrect and correct alternatives.
- Avoid the “all of the above” option-it can penalize slow readers and students who select a correct option without reading further; it can reward a student for incomplete understanding (if 2 options are correct, “all of the above” must be the answer).
- Good distractors let you know where students went wrong-create an incorrect answer to match each type of error a student could make.
- Use distractors that could possibly be correct-adding an extra distractor that is obviously incorrect does not improve an item (remember that all items do not have to have four or five distractors) .

This is a long list of recommendations-one reason for its length is that multiple-choice questions are one of the most frequently used test item format. Educational

Assessing Student Learning

researchers have devoted a great deal of time and effort to studying this format. Although the list is long, you should remember that it is difficult to follow all of these guidelines all the time. Maximizing good testing practice and minimizing difficult or confusing items are the goals you should strive to attain.

Writing Good Essay Questions

One reason that faculty may turn to using essay questions for assessment is that they can write essay questions in a minimum amount of time, particularly compared to good multiple-choice questions. However, as many faculty have later discovered, an essay question written in haste may turn out to be quite difficult for students to answer and/or hard to grade easily or well. One way to avoid these problems is by following guidelines developed to facilitate writing good essay questions. Hales and Marshall (2004, pp. 159-165) listed nine guidelines for developing quality essay items.

- Give yourself adequate time to write questions-as I mentioned previously, a hastily written item may turn out to be a bad one; you should write your questions and give yourself time to evaluate them before using them.
- Use learning objectives to guide your question writing-just as with multiple-choice questions, your essay questions should be firmly rooted in your learning objectives.
- Give students a problem to solve in each question-you can use multiple-choice questions to assess knowledge of information; an essay question should ask students to use their information in some way-usually higher order thinking skills.
- Define the problem clearly-you want all your students responding to the same essay question; if they have to interpret it, you will likely get multiple interpretations.
- Keep the problem limited-although essay questions are usually broader than multiple-choice questions, a question that is too broad tends to overwhelm students because of a lack of direction.
- Give explicit directions-avoid making the students guess “what does the teacher want?” If you want examples, complete sentences, a graph, or have some other specific requirement, say so.
- Avoid optional questions-when you use optional questions, students end up taking different exams. Students of different abilities may select different

questions; the instructor might react more favorably to some questions than to others. Also, if students know they will have a choice of essays, they may choose not to learn some of the material.

- Make a scoring guide for each question-even if you do not write it, you should have a “best answer” in mind. Preparing a scoring guide will minimize scoring errors and increase the reliability of grading.
- Use a structured scoring process to minimize scoring errors-scoring essay answers without knowing the writer will minimize bias, whether positive or negative. Grading all of the same essay for all students before moving to another essay will help you keep the scoring guide in mind.

Using Rubrics for Scoring Essay Answers

A rubric is a standardized scoring guide for an essay answer. Rubrics are often used when there are different people scoring the same question over many students-for example, scoring the free response section of Advanced Placement exams in the US, which can have hundreds of graders scoring thousands of student responses. A team of experienced question readers develops a rubric and then trains other question readers in using the rubric so that all students' essays are scored using the same criteria, even though there are many different readers.

To create a rubric, you must first decide all the different dimensions or elements on which you want to score an essay answer (or any student performance such as a term paper, classroom presentation, etc.). For example, Hales and Marshall (2004, p. 203) gave an example rubric used for grading an essay for an English composition class. The instructor decided to score students on their Ideas and Content, Organization, Voice, Word Choice, Sentence Fluency, and Convention (e.g., spelling, punctuation). At that point, the instructor had to decide whether to count those six criteria equally or weight them differentially-because it was a composition course, the instructor counted the criteria equally. Finally-and this point is key in building a rubric-the instructor must develop a point system for rating each of the criteria and give a verbal example. In this example, the instructor decided to use a 3-point rating scale for each criterion. For the Ideas and Content criterion, the instructor assigned 3 points for essays with “Clear main theme; strong ideas; high-level detail” (Hales & Marshall, 2004, p. 203). An essay judged to have “A discernable main theme not clearly articulated; insufficient detail” (Hales & Marshall, 2004, p. 203) received 2 points for Ideas and Content. Finally, a student received 1 point for the Ideas and Content criterion if the

Assessing Student Learning

instructor found “No main theme; little detail” (Hales & Marshall, 2004, p. 203). In this manner, the instructor could grade each student’s essay on these six criteria, with each criterion being scored on a 1-3 basis. Thus, an excellent essay would receive a score of 18 (6 criteria x 3 points each), whereas a very weak essay would be scored with a 6 (6 criteria x 1 point each). Using a rubric makes the scoring a simple matter for the instructor. By giving each student a copy of the rubric with marked scores, the instructor is able to provide much more detailed feedback than simply giving an overall grade. Both of these outcomes are ideal as far as assessment is concerned.

Summary

Clearly, assessing student learning well is much more involved than some faculty and (probably) most students believe. However, given the crucial nature of faculty members’ jobs of assessing learning, it is important to take the time necessary to do a good job. Note how much of this booklet has been devoted to important aspects of assessment *before* actually writing the assessment instrument. I hope that this pattern sufficiently emphasizes the importance of being well prepared to assess student learning before you write an exam.

In closing, the important points to take away from this booklet include the following:

- Class assessments can be either formative or summative.
- Formative assessment is designed for the purpose of providing students with feedback about their learning, with no implications for a grade.
- Summative assessment involves assessing student learning for grading purposes.
- High-stakes assessment uses a test to solely or largely determine ultimate student success. High-stakes assessments are summative in nature.
- Low-stakes assessment plays no role or only a small role in students reaching a goal. Low-stakes assessments are typically formative in nature, although a summative graded assignment such as a quiz that weighs little in a student’s grade could be low-stakes.
- Assessing student learning is linked to developing student learning outcomes, which follow from your learning goals.
- Student learning outcomes must be stated in measurable terms.
- Learning outcomes are typically linked to a level or type of student learning.

Assessing Student Learning

- An objective assessment is essentially any test that can be computer scored, such as multiple-choice, true-false, or matching questions, plus fill-in-the-blank questions.
- Nonobjective assessments involve tasks such as writing, research, or some other task completion.
- Learner-centered assessments aim at modifying behaviors of the learners rather than the instructor. Such assessments tend to be formative in nature.
- To develop an assessment instrument, you should begin with a test blueprint—an outline for your test that includes all the learning objectives that you want students to know for the test.
- Research has documented best practices for writing both objective items and nonobjective items for assessment.
- A rubric is a standardized scoring guide for an essay answer. Using a rubric to grade nonobjective items allows for scoring to be more uniform and reliable.

References

- Abbott, R. D., & Falstrom, P. (1977). Frequent testing and personalized systems of instruction. *Contemporary Educational Psychology*, 2, 251-257.
- Frequent testing in a lecture course led to higher achievement than less frequent testing.
- Anderson, L. W., & Krathwohl, D. (Eds.). (2001). A taxonomy for learning, teaching and assessing: A revision of Bloom's taxonomy of educational objectives. New York, NY: Longman.
- An updated version of Bloom's (1956) taxonomy of educational objectives; seems to be much more commonly used and cited now than Bloom's version.
- Angelo, T. A., & Cross, K. P. (1993). *Classroom assessment techniques: A handbook for college teachers* (2nd ed.). San Francisco, CA: Jossey-Bass.
- The best source for quick, formative classroom assessments of student learning and understanding that you can find.
- Bangert-Drowns, R. L., Kulik, J. A., & Kulik, C.-L. C. (1991). Effects of frequent classroom testing. *Journal of Educational Research*, 85, 89-99.
- Students who were frequently tested scored about 0.1 standard deviations

Assessing Student Learning

higher than less frequently tested students.

- Bloom, B. S. (1956). *Taxonomy of educational objectives: The classification of educational goals*. New York, NY: Longmans, Green.
- This book is the standard by which all learning-based taxonomies are judged; this book began the focus on student learning rather than teacher performance.
- Cangelosi, J. S. (2000). *Assessment strategies for monitoring student learning*. New York, NY: Addison Wesley Longman.
- A guide to help teachers make complex instructional decisions on the basis of assessing student learning.
- Churches, A. (2008). Bloom's taxonomy blooms digitally. *Tech&Learning*, <http://www.techlearning.com/article/8670> Churches developed an updated version of Bloom's (1956) taxonomy that focused on student learning through digital formats.
- Hales, L. W., & Marshall, J. C. (2004). *Developing effective assessments to improve teaching and learning*. Norwood, MA: Christopher-Gordon.
- This book is focused on improving student learning through assessment and data-based decision making.
- Peckham, P. D., & Roe, M. D. (1977). The effects of frequent testing. *Journal of Research & Development in Education*, 10, 40-50.
- Reviews the evidence from studies of frequent testing.
- Smith, R. A. (2011). *Writing student learning objectives*. Deanship of Skills Development Booklet Series, King Saud University.
- This booklet in this series examines the practice and advantages of writing student learning objectives.
- Suskie, L. (2009). *Assessing student learning: A common sense guide* (2nd ed.). San Francisco, CA: Jossey-Bass.
- A comprehensive guide to assessment, including understanding assessment, planning for assessment, developing assessment instruments, and using assessment results.

© King Saud University, 2013

King Fahd National Library Cataloging-in- Publication Data

L.D. no. 1434/ 7296

ISBN: 978- 603- 507- 128- 4



عمادة تطوير المهارات
Deanship of Skills Development

King Saud University - Deanship of Skills Development

P.O. Box 85500 Riyadh 11691

d s d . k s u . e d u . s a