

Using Exams for Measuring Learning Gains

Barbara Glesner Fines
University of Missouri–Kansas City School of Law

The title of this session is measuring “learning gains.” It is an intriguing title when one thinks about what exams do. Since few professors obtain baseline data through pre-testing of skills, knowledge and values, most exams in courses test only “learning” --- assuming that the “gains” in learning have occurred in the course. There is much to be gained by considering the use of diagnostic and formative assessment. Practice exams and other formative assessment (minute papers, quizzes, etc.) can be used quite efficiently for this purpose and to reinforce learning while there is still time for student correction. If faculty members shared their assessment data across courses, summative assessments from one course could be used as baseline data for the next.

This does not mean sharing student grades. Probably the single greatest confusion in the vocabulary of assessment is that which equates assessment with Grading. It is easy to confuse the language of assessment and the language of grading and to presume that, because a law teacher grades, he or she is assessing to improve student learning. Indeed, the Best Practices chapter on assessment began with a discussion of grading.¹ However grading and assessment are different in critical ways. First the purposes differ: grading is evaluative -- designed to sort and classify students, to confer distinction or channel students; assessment is ameliorative -- designed to improve student learning and faculty teaching. Second, grades rarely give law schools the kind of information that can lead to that improvement, as grades do not reference explicitly any clear criteria.² Thus, assessment requires measurement rather than simply “grading.”

What does measurement require?

Simply put, concrete and comparable data. This is information that is usually translated into numbers or ratings (even letter grades, so long as you know what the grade means in terms of learning outcomes!). However, this concept of measureable learning outcomes does not require a statistical precision – good, better, and best are forms of measurement that, so long as the rater applies the same criteria consistently, can provide plenty of information about student learning and permit comparisons across time or classes. Whether 1, 2, or 3 or Good, Better, Best, the critical feature for assessments is to be able to measure student learning in a way that provides useful information. Numbers simply have the advantage of being easily manipulated into averages, standard deviations, etc. Any qualitative measure can be translated into some numerical value for these purposes, but for purposes of consistency in measuring performance, qualitative labels are often more helpful in providing benchmarks.

While outcomes assessment might use the very same methods (tests, essays, papers, performances) as those used to assign grades, assessment requires more analysis of the components of those grades. So, for example, knowing that ten percent of students received an A on an exam or a clinical performance, tells a teacher almost nothing about what the students learned or where to focus efforts to improve learning. Suppose, however, law teachers have defined the criteria by which they are judging the students’ work, describing varying levels of competence on defined subsets of knowledge, skills, or values (this would be referred to as a Rubric in educational literature). These law teachers could then

¹ ROY STUCKEY AND OTHERS, BEST PRACTICES FOR LEGAL EDUCATION 235 (2007).

² Best Practices emphasized this point in suggesting that grading be criterion referenced. *Id.* at 243-45.

gather information about student learning that they could use in the future to improve overall learning. If law teachers agree upon common rubrics or criteria for assessment in their individual classes, data can be aggregated to gain an overall picture of student learning, growth, and deficiency. For comparative purposes shared questions (across courses or across semesters) and shared rubrics, permit the professor to exam whether changes in the learning activities in the course result in changes in learning gains.

Using Exams --What kinds of exams best measure learning gains?

The simple answer is “Whatever kind measures the learning outcomes you are aiming for.” If a professor is testing recognition of information, multiple choice/ short answer questions are efficient and will accurately capture the student learning. However, most professors would say that they are aiming higher on Bloom’s taxonomy to test application and synthesis of knowledge. For these outcomes, one can still use multiple choice but it will be more difficult to see the student’s thought process. Guessing is invisible. However, very well constructed multiple choice questions can uncover misconceptions and errors in reasoning quite well. To construct these question effectively, the professor must be able to know what those misconceptions are. Here is where using prior exams and formative assessments permits refinements in future assessments.³ Essay questions allow finer diagnosis of confusions and incomplete understandings.

One of the most difficult challenges in designing examinations that test knowledge and analytical skills is keeping the focus on the skill being tested. For example, if a professor says “I teach statutory construction” but an examination question requires students only to apply the knowledge contained in statutes that they have studied during the semester, the students may not truly be demonstrating their ability to read and interpret a statute. To more directly measure this skill, the exam should provide students with a new statute or rule to read and interpret. As to other analytical skills as well, more precision about what those precise objectives are will permit better targeting of measurement of the skill.

In testing skills, one objection is often that measurement of skills is fraught with subjectivity. As *Best Practices in Legal Education* noted, there may be some important learning goals that can’t be quantified and so cannot be measured.⁴ However, many leaders in legal education took to heart the urging of Best Practices that “if law teachers make the effort, we may discover ways to evaluate some things that we might initially consider unmeasurable” and that “We should not stop trying to achieve desirable outcomes because they are difficult to assess.”⁵ In many cases, what law teachers have discovered that the difficulties are not found in the nature of assessment, but in the degree to which law teachers are able to describe succinctly and concretely the competencies they are targeting.⁶

Another objection is that testing of skills will often involve different experiences from student to student, and so are not valid measures. Assessment of student learning on disparate tasks is not invalid

³ See “The Wrong Answers are the Most Important Questions” in Barbara Glesner Fines, *Lessons Learned About Classroom Teaching from Authoring Computer-Assisted Instruction Lessons*, 38 WILLIAM MITCHELL L. REV. 1094, 1112-1115 (2012), available at Available at: <http://open.wmitchell.edu/wmlr/vol38/iss3/2>.

⁴ Best Practices at 253.

⁵ *Id.*

⁶ For examples of final exams using skills-based exercises such a client interviews, team-based learning assessments, reflective essays, and others, see Barbara Glesner Fines, *The Power of a Destination: How Assessment of Clear and Measurable Learning Outcomes Drives Student Learning* (2014) at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2529086.

so long as the professor has clearly defined the knowledge, skills, or attitudes that the students are expected to demonstrate in performing those tasks and so long as clear criteria are used to measure that performance along a spectrum of clearly defined levels of proficiency. Our belief that all students must perform the same activity to assess their performance is grounded more in our concerns for “fair grading” than it is in concerns for discovering meaningful information about student learning. Shared development of rubrics can provide consistency across multiple graders.

Of course, the purpose of all this effort is to improve student learning. Which takes us full circle to the concept of “learning gains.” If we think of teaching as “adding value” to the students’ stock of knowledge, skills, and values, assessment is the tool by which we can measure and improve that value.

Assessment Tales: The Bluebooks That Stayed

Barbara Glesner Fines

January 17, 2012

<http://bestpracticeslegaled.albanylawblogs.org/2012/01/17/assessment-tales-the-bluebooks-that-stayed/>

It’s that time of year when we all have the grading of our last semester’s bluebooks well behind us and the last few students have come in to review their exams. So we have packed up the bluebooks to be archived and they are out of sight and out of mind.

But wait! In the world of assessment, the bluebooks come back! These papers and exams have a wealth of assessment information for us to mine if we only take the time to gather, reflect and use that information.

How can you use your bluebooks for efficient assessment and improvement of student learning?

Many faculty gather holistic impressions as they grade about the performance of the students overall and the areas of difficulty and strength. To improve on this reflective process, faculty can take a few more simple steps:

1. Improve your data collection.

Rather than gathering general impressions as we grade bluebooks, we can mine the bluebooks for some more concrete data. Examine the distribution of performance on individual questions or issues. Note that you need not gather every data point possible from the bluebooks. Often it is helpful to begin with two or three items to analyze. For example, what is the one thing that nearly every student did well on the exam? What were the one or two questions/issues/approaches that many students had problems on? What percentage of the students had these problems?

2. Analyze your data.

For issues students appear to have learned well, look again at your questions. How confident are you that the question truly tested the student understanding? In this respect, essay questions are often easier to evaluate than multiple choice questions, because you can see the students reasoning on the former, whereas consistently correct answers on the latter can be the result of distractors that are patently wrong. What materials and techniques did you use to prepare the students for that question? When during the semester did you teach those matters? If the student performance is improved from prior exam administrations, what, if anything, did you change that may have caused this improved learning?

For issues or questions on which a significant percentage of student performance was deficient, again, begin by reexamining the question, its placement in the examination and the time allocated for responses, to identify other possible reasons for poor performance that are less related to student learning and more related to exam conditions. Look for patterns in the student errors or misconceptions that can help you diagnose what learning conditions led to the student poor performance. What materials and methods did you use to teach this doctrine?

3. Plan for the next class

When students are performing well on a doctrine or concept, especially when that competent performance appears to have been the result of your prior efforts to target and improve learning activities for that material, you may be tempted to rest on your (and your students') laurels. However, consider that any change to one part of a course can affect other parts and each class brings with it different experiences and preparation.

To improve student learning on areas that have presented difficulties for students, consider not only improving teaching materials or methods related to that area, but also incorporate more formative assessments during the term to help you and the students identify earlier and more clearly the learning deficiencies.

4. What my bluebooks told me this semester:

To illustrate this process of mining bluebooks for assessment, I will discuss this semester's Professional Responsibility exam. From this semester's bluebooks, I gathered a range of data on materials well understood and poorly understood. I will share three examples of data to illustrate the process of using bluebooks for an assessment process.

The doctrinal winner this year in terms of student performance was multijurisdictional practice of law. Is this because the students understood these aspects of the course better than others? Reviewing the exam, I noticed that the question testing this subject called for a fairly low level of mastery (basic issue spotting and knowledge of rule) without any sophisticated analysis required. This was a topic for which I had provided a number of practice problems to the students and I had tested the issue in a similar fashion on a prior year's exam, which I had made available for student review. Moreover, it is a subject that, because my law school is located on a state line, with dramatically different variations on this rule, the students understood that this was a rule that would impact their immediate future, as they chose which state bar exam to take first. What I learned from this is the fairly unremarkable understanding that my law students can and will

master at a knowledge-level those topics for which they know they will be tested and for which they also have a more personal motivation to learn well. I concluded that I would and could generalize these understandings to not only raise the bar on testing this doctrine, requiring a more sophisticated understanding, but also would look for other areas in which I could improve student motivation by identifying the specific need-to-know circumstances looming in their immediate future for other rules.

A second topic about which I have been tracking student learning performance for many semesters is the student understanding of the distinction between the evidentiary attorney-client privilege and the ethical duty of confidentiality (among other doctrine). When I first began tracking, as many as 30% of students were demonstrating fundamental confusion on this topic – using language of “privilege” when the subject was confidentiality (or vice versa) or confusing the exceptions to the ethical duty with the crime-fraud exception to privilege. I knew from speaking with other Professional Responsibility teachers that this is a common area of confusion for students. Over the course of several semesters, I worked to improve student learning in this area: including more problems in course materials, writing and assigning a CALI lesson on the subject, and explicitly telling the students that this is something that I am tracking and cheering them on to “make this the 100% mastery year.” The efforts are bearing fruit. This semester was the best yet – only four out of 72 students used the vocabulary of the two doctrines improperly and three of these applied the correct rule even though they were not using the correct terminology in doing so.

An area on which I had thought I was making progress in student learning turned out to be a continuing problem. Students commonly are confused by the rule governing an attorney’s right to withdraw from representation. I have made the same efforts on this doctrine as I have with the privilege v. confidentiality confusions: increasing problems, providing additional outside resources (again, I wrote a CALI lesson on the subject); and providing in-class quizzes to assess understandings while there was still time to improve learning. However, I was puzzled to see 13 of the students declare that an attorney may not withdraw from representation if it would harm the client. What could have been the source of this confusion? Searching through my course materials and lesson plans, I uncovered the problem. A powerpoint lecture on withdrawal from representation when the client fails to pay the attorney contained a page with a bulletpoint list of reasons that courts might deny an attorney permission to withdraw even though the rules would permit the withdrawal. One of the bullet points listed “degree of harm to the client” as a factor the court would consider. Obviously some students had transferred the powerpoint slide into their notes on the general withdrawal rule rather than recognize that these factors were connected only to the judicial discretion to deny an otherwise permissible withdrawal. Again, a well-worn lesson learned anew: as helpful as powerpoint slides can be for organizing discussions and providing visual cues for learning, students will study text of these slides as definitive statements of law rather than thumbnails of larger discussions and understandings. Conclusion: no shortcut summary slides!