

Assessing Student Learning: Strategies and Lessons from the Library Instruction Classroom

Kate M. Manuel
University Library
New Mexico State University
Las Cruces, NM 88003
kmanuel@lib.nmsu.edu

Abstract

Program evaluation “involves a purposeful appraisal or check on the effectiveness and efficiency of the various components of the instructional program” and “is . . . key to discovering when to improve presentations, revise materials, and modify teaching methods” {11}. Assessment is a specific type of evaluation focusing on students’ ability or achievement. Assessment has recently become important in education as a means of enforcing standards and promoting accountability {13}, and library instruction – probably because of its relatively recent entry into the higher education curriculum – has been particularly asked to “provide concrete proof of its usefulness” {12}. As a result, this author has had to master a range of methods for assessing student learning – from “traditional” pre- and post-testing to performance, portfolio, and other “alternative assessment” methods relying upon rubrics. This paper provides an overview of the major types of assessment instruments, methods for applying them to classroom settings, and the types of feedback that each can provide. Samples of various assessment instruments from a range of disciplines are provided for consideration.

Assessment: An Overview

In a March 1996 article in *Education*, Angela Maynard Sewall suggestively summarized the recent history of education by writing

... no doubt future historians will view the 1980s as the education decade – not a decade in which specific education problems were solved, but a decade in which the problems of education as a whole achieved national significance. If that is true, then the decade of the 1990s may well be known by historians as the decade in which national pressure was brought by non-educators to cause educators to change practice and theory. The 1990s will be known as the decade of accountability. This will be the case not because of the imposition of standards but because of the process and practice of assessment which increasingly has become the focus of legislators, politicians, and the public during the past five to ten years {21}.

While the developments of the next seven years cannot reliably be predicted, it seems not implausible that the first decade of the 21st century will see the continuation and perhaps culmination of the interest in assessment that characterized the 1990s. Certainly the unhappiness of some politicians, pundits, and members of the general public with a higher education system that is perceived as unproductive by market standards and unresponsive to changing societal needs continues to be expressed.¹ While assessment should not, ideally, be undertaken simply to meet the demands of external bodies for educational institution’s accountability, responsiveness to professional and regional accreditation agencies, as well as to state legislatures and taxpayers, remains an impetus for many assessment activities {7, 22}. A “purer” reason for assessment is situated by Peggy L. Maki, Director of Assessment for the American Association for Higher Education, in “institutional curiosity,” “a desire to discover the fit between institutional or

¹ To take but a few examples, *Forbes* magazine has recently taken education to task because “[a]s education has grown to compete with medical care for the claim of being America’s largest single industry, it has so far resisted all efforts to improve its productivity” {18}, while a recent *El Paso Times* editorial page story characterizes the pay of university presidents as “wretched excess”{16}.

programmatic expectations for student achievement and patterns of actual student achievement” {17}. Assessment here ties to maximizing student growth, improving instruction, recognizing (teachers’ and/or students’) accomplishments, and modifying and improving instructional programs {14, 26}.

Despite being simply definable as the “process of collecting, analyzing, and reporting data” {14}, assessment remains a touchy issue for many educators. The movement toward classroom assessment techniques (CATS) led by Thomas Angelo and Patricia Cross – which has familiarized higher education with many wonderful methods, such as the one-minute paper and RSQC2 (recall, summarize, question, comment, and connect) for measuring student learning in the classroom setting – is largely an outgrowth of some educators’ fears that national standards would be applied in assessment efforts and used to “prove” that they were “bad” teachers. National, standardized assessment instruments, which will be discussed later in this paper, loom particularly large in negative reactions to assessment.² There is concern that because “accountability is facilitated if schools, classrooms, and students can be compared” {8}, overzealous administrators may implement national, standardized tests without consideration of potential mismatches between the tests and the particular programs being evaluated {15}. Many instructors worry that although “assessment can be seen as part of the teaching process itself rather than as a separate task” {14}, it results in shifting the “emphasis away from what [some consider] the most important element of student learning: student interaction with professors in a classroom” {22}.³ Then there is the complication that “[a]ssessment is done to satisfy the requirements of various audiences whose needs are changing” {23} – meaning that projects must juggle the interests of students, instructors, and others in the results of the assessment while these interests are themselves in the process of changing (e.g., one year, the public’s focus might be on scientific literacy and the next year on phonics). Then there are the facts that assessment is an on-going, iterative process {14}; that it often requires agreement on what to assess (learning outcomes) {17}; and that it yields data a lot “fuzzier” than most of the data with which academic researchers are used to dealing.⁴ Finally, overly ambitious assessment projects – or projects whose results seem to no further than some administrator’s filing cabinet⁵ – often sour participants on the entire project of assessment {17}. The author’s own profession of librarianship has been particularly prone to “grand” assessment projects lately: ever since the Association of College and Research Libraries’ adoption of the *Information Literacy Competency Standards for Higher Education* in 2000, librarians have been hard at work devising a single assessment instrument that measures all 22 performance indicators and 88 outcomes simultaneously. It has yet to work.

Major Types of Assessment Instruments, Methods of Application, and Implications of Each

As Janet Williams of the Educational Testing Service suggests, “Assessment is not a mystery. In fact, it is something we do multiple times a day, every day. For example, one may step on the scale in the morning to weigh oneself” {25}.⁶ Distinguishing between the various types of assessment instruments that exist and selecting the one that best meets one’s needs, however, can seem mysterious. [See Tables 1 and 2 for an overview of various types of assessment instruments.] There are **psychometric** instruments, designed to measure respondents’ knowledge, aptitude, ability, etc.; **sociological** ones, measuring respondents’ attitudes or values; and **illuminative** ones (also called **goal-free** or **naturalistic**), based not on pre-set objectives but rather focusing on the effectiveness of programs, services, etc. in meeting the aims and needs of the community as they are defined by the community. Psychometric and sociological instruments can be further described as either **selected-response** or **constructed-response**, depending

² One major objection to national, standardized tests is that they reify existing socio-economic differences among students {20}.

³ Eisner {8} suggests that “[u]nder optimal teaching, variability [between students] would surely increase.” Others claim an inherent “conflict between accountability and the goals of instruction” {13}.

⁴ Control groups, for example, are generally lacking within the context of educational assessment {10}.

⁵ “Assessment is certain to fail if an institution does not develop channels that communicate assessment interpretations and proposed changes to its centers of institutional decision making” {17}.

⁶ As Williams {25} notes, “Assessment, measurement, and testing are used interchangeably.”

upon whether those tested select their answers from among various choices provided on/by the test (i.e., selected-response or **structured**) or whether those tested compose answers to questions *de novo* (i.e., constructed-response or **non-structured**). Both selected- and constructed-response instruments can themselves be either **criterion-referenced**, measuring how well examinees perform relative to specified standards or criteria, or **norm-referenced**, measuring how well examinees perform relative to other examinees. Some constructed-response instruments (e.g., work simulations or work sample tests) are **authentic assessment** instruments because they measure performance on “real world” tasks and situations. Assessments can also be **formative**, looking at progress during a span of instruction, or **summative**, looking at the end results of an instructional span. These types are not mutually exclusive; any one instrument can be of multiple types.

Summative assessments, such as end-of-term student examinations or teacher-course evaluations, or inventorying of students’ abilities at the completion of a degree program, are commonly used, but formative assessments are more helpful to instructors and programs by allowing them to measure progress in student learning while there is still time for further or revised instructional interventions. When the only form of assessment used is summative, special importance attaches to **pre-testing**: without some initial measure of students’ skills and knowledge, there is no way of knowing for sure whether the learning outcomes claimed for the instructional unit are, in fact, the results of this instructional unit, or whether they pre-existed it {2}. Any of the following types of instruments can be used in either formative or summative assessment projects.

Assessments can also be either selected response or constructed response. Selected response instruments require students to select their answers from among the options given; examples include true/false, multiple choice, matching, checklist, and rank ordering questions. Selected response instruments are commonly faulted for the inability to test higher-order thinking skills,⁷ but they are much easier to administer and score than constructed response instruments. **Constructed response** instruments require students to supply their entire answer from their own knowledge and in their own words. They are more easily suited to assessing higher-order thinking skills {6}. Examples range from short, fill-in-the-blank questions to lengthy essays. Extracting meaningful data from constructed response instruments can be tricky, though; machine scoring of answers is far from perfect, human scoring takes substantial time, and subjective factors quickly intrude into scoring.⁸

In this, constructed response instruments resemble the so-called “alternative assessment” or “performance assessment” instruments that currently hold much cachet within education. Such approaches “aim[at moving away from testing practices that require students to select the single correct answer from an array of four or five distractors to a practice that requires students to create evidence through performance that will enable assessors to make valid judgments about ‘what they know and can do’ in situations that matter”{8}. Examples include portfolios or exhibitions of student work, performance of laboratory experiments, students’ responses to simulations, observations of students performance(s) on a task, etc. Ideally, these instruments give a much better sense of students’ knowledge by asking them to respond to “real” situations in a problem domain, rather than regurgitate random facts memorized from lectures or textbooks. The problem this author has found with some such instruments is that while more “real” to students’ actual problem solving situations than multiple choice tests, they still seem “inauthentic” to students and do not necessarily motivate them to put forth their best efforts. Portfolios submitted at the conclusion of coursework are a standard method of assessing student learning in information literacy and library research courses. A portfolio, in theory, provides “documentation of student growth based on the student’s learning goals established as the beginning of the portfolio completion” and thus “must contain deliberate compilations, not casual collections, or items” {14}. In practice, however, portfolio keeping is not natural to the information research process (in real life most

⁷ Another problem of selected response instruments is that the “content or format of multiple-choice tests unfairly favors students from traditionally higher-scoring groups”{13}.

⁸ As Stripling {24} writes, “Writing is the most developed form of performance assessment; with extensive training, evaluators for written work have finally achieved some reliability.”

people gather information without making notes of where or how they searched, stopping when they are satisfied, not when they have examined a specified number of resources), and a number of people cannot muster a “realistic” response to a “contrived” research situation. Performance measures have proven equally problematic. Either students must record their own responses in writing or orally for future review, and/or screen capture software or observations must record their behaviors. The problems here, of course, are that proficient information seeking strategies do not always correspond to articulate expressions of them,⁹ and tracking student behaviors without benefit of student comments can lead to mistaken inferences {1}.

Lengthy constructed response instruments, alternative assessments, and performance assessments are typically “scored” in relation to benchmarks or **rubrics**. Benchmarks or rubrics codify learning outcomes in the form of content and/or performance standards; they described students’ performances at various levels of proficiency {9} for example, by degrees of quality, frequency, expertise, or other measures {4}.¹⁰ While scoring rubrics do increase the reliability and reliability of such instruments, it still remains difficult to compare performances to external standards {24}. Reliability of scores is particularly an issue, as different graders may rate the same work differently, or the same grader could rate the same work differently at different times {6}.

Reliance on national, standardized assessment instruments relieves many of these worries about reliability and validity. The testing experts producing these instruments can, presumably, be counted on to create ones with solid reliability and validity. National, standardized instruments also come with instructions regarding administration of the instrument that are, perhaps, more likely to be abided by than local instructions. “Good assessments are not easy to create” {25}, and most academic faculty outside of education departments have little expertise or interest in testing issues. As the “validity of test-based data is entirely dependent on the quality of the test” {2}, one can see the impetus to rely upon tests of known reliability and validity rather than locally produced ones. The major problem, through, is that national, standardized instruments cannot measure uniquely local outcomes.¹¹

Conclusions

All assessment activities should have a goal or purpose and a projected use for the data gathered therefrom: as the Educational Testing Services’ Williams {25} cautions, “know how you are going to use the information before you design the instrument.” It is important to keep this in mind – as sometimes it seems as if the impetus to assess, for the sake of assessing, prevails over the more difficult task of determining what is to be assessed and how. Assessment should not be an end in itself, and there is need to resist the one-upmanship sometimes displayed in assessment circles. Some people seem to be more interested in proving that they assessed the largest sample population, or used the most alternative measures, or measured the largest range of outcomes than in applying the results of their data to improve their programs. This is, ultimately, the purpose of assessment.

Works Cited

1. Adams, M. & Morris, J. M. (1985). *Teaching library skills for academic credit*. Phoenix, AZ: Oryx.
2. Barclay, D. (1993). Evaluating library instruction: Doing the best you can with what you have. *RQ*, 33 (3), 195-203.

⁹ While some would say that “the true measure of learning is how well knowledge can be expressed in writing” (see the Knowledge Analysis Technologies Web site at <http://www.knowledge-technologies.com/content.html>), others favor application of knowledge – not articulation of it – as its true measure (cf. {3}). As Catts {6} notes, “Theory papers are prone to favor the literate and especially those with a broad general knowledge. Practical test that ignore theory and understanding can also be invalid.”

¹⁰ With rubrics one should use, “language that describes rather than labels performance.” {14}.

¹¹ Another problem with them is that while the “differential effects of tests on people from diverse social groups are not limited to standardised tests, the problem is of particular significance for such tests” {6}.

3. Berman, E. (2000). *Using pretests*. Available at http://aer.arizona.edu/Assessment/New_Short_Guide/using_pretests.htm.
4. *Building rubrics*. (No Date). Available at <http://www.sover.net/~mttop/arts/rubrics>.
5. Campbell, D. (2000). Authentic assessment and authentic standards. *Phi Delta Kappan*, 81 (5), 405+.
6. Catts, R. (2000). Some issues in assessing information literacy. In *Information literacy around the world* (pp. 271-283). Wagga Wagga, New South Wales: Centre for Information Studies.
7. Dunn, K. (2002). Assessing information literacy skills in the California State University: A progress report. *Journal of Academic Librarianship*, 28 (1), 26-35.
8. Eisner, E. (1999). The uses and limits of performance assessment. *Phi Delta Kappan*, 80 (9), 658+.
9. *FAQ's: Frequently asked questions about alternative assessment*. Available at <http://www.oise.on.ca/~ascarfone/faq.htm>.
10. Fenske, R F. (1998). Computer literacy and the library. *Reference Services Review*, 26 (2), 67-78.
11. Freedman, J. & Bently, H. (1986). Techniques of program evaluation. In *Teaching Librarians to Teach* (pp. 188-204). Metuchen, NJ: Scarecrow Press.
12. Geffert, B & Bruce, R. (1997). Whither BI? *RQ*, 36, 409-422.
13. Haertel, E. (1999). Performance assessment and education reform. *Phi Delta Kappan* 80 (9), 662+.
14. *Information power: Building partnerships for learning*. (1998). Chicago: American Library Assoc.
15. Kaplowitz, J. (1986). A pre- and post-test evaluation of the English 3-library instruction program at UCLA. *Research Strategies*, 4, 11-17.
16. Lockman, N. A. (2002, December 1). Many university presidents boarding the gravy train. *El Paso Times*, pg. 15B.
17. Maki, P. L. (2002). Developing an assessment plan to learn about student learning. *Journal of Academic Librarianship*, 28 (1), 8-13.
18. Michaels, J. W. & Smillie, D. (2000, May 15). Webucation. *Forbes*, 165 (11), 92-94.
19. Moorcroft, T. A. et al. (1999). Authentic assessment in the informal setting: How it can work for you. *Journal of Environmental Education*, 31 (3), 20+.
20. Ramirez, A. (1999). Assessment-driven reform: The emperor still has no clothes. *Phi Delta Kappan*, 81 (3), 204+.
21. Sewall, A. M. (1996, March 1). From the importance of education in the 80's to accountability in the 90's. *Education*, 116 (3), 325-329.
22. Stallings, D. (2002). Measuring success in the virtual university. *Journal of Academic Librarianship*, 28 (1), 47-53.
23. St. Clair, G. (1995). Assessment: How and why. In *Academic Collections in a Changing Environment* (pp. 58-70). Westport, CT: Greenwood Press.
24. Stripling, B. K. (1994). Assessment of student performance. In *Assessment and the School Library Media Center* (pp. 77-93). Englewood, CO: Libraries Unlimited.
25. Williams, J. L. (2000). Creativity in assessment of library instruction. *Reference Services Review*, 28 (4), 323-334.
26. Young, R. M. & Harmony, S. (1999). Working with faculty to design undergraduate information literacy programs. New York: Neal-Schuman.

Author's Biography

Kate Manuel is currently Instruction Coordinator for the New Mexico State University Library. She was previously Physical Sciences Instruction/Reference Librarian at California State University, Hayward. She has spoken and written extensively on assessment and library instruction, including presentations at the 2002 New Mexico Library Association Conference and at the 2001 Library Orientation Exchange Conference and publications in *Managing Instruction Programs in Academic Libraries* (Pierian Press), *Library Hi-Tech News*, *College and Research Libraries News*, and *Issues in Science and Technology Librarianship*.

Table 1: An Overview of Major Types of Assessment Instruments

Types of Testing Instruments	Selected-Response Measures <i>Those tested select their answer from among various choices provides</i>		Constructed-Response Measures <i>Those tested compose an answer to a question</i>		
<p>Psychometric <i>Designed to measure respondents' knowledge, aptitude, etc.</i></p>	<p>Criterion-Referenced <i>Measures how examinees perform relative to other examinees</i></p> <p>Multiple Choice Tests</p> <p>True/False Tests</p> <p>Matching Items</p>	<p>Norm-Referenced <i>Measures how examinees perform relative to specified standards or criteria of performance</i></p> <p>Multiple Choice Tests</p> <p>True/False Tests</p> <p>Matching Items</p>		<p>Criterion-Referenced <i>Measures how examinees perform relative to other examinees</i></p>	<p>Norm-Referenced <i>Measures how examinees perform relative to specified standards or criteria of performance</i></p>
			<p>“Inauthentic” <i>Based on classroom or academic tasks and situations</i></p>	<p>Essay Questions</p> <p>Problem Solving</p> <p>Oral presentations</p> <p>Portfolios</p> <p>Simulations</p>	<p>Essay Questions</p> <p>Problem Solving</p> <p>Oral presentations</p> <p>Portfolios</p> <p>Simulations</p>
			<p>Authentic <i>Based on “real world” tasks and situations</i></p>	<p>Work Simulations</p> <p>Work Sample Tests</p>	<p>Work Simulations</p> <p>Work Sample Tests</p>
<p>Sociological <i>Designed to measure respondents' attitude, values</i></p>	<p>Surveys/Questionnaires →</p> <p>Self-report Measures →</p> <p style="text-align: center;">With scales, checklists, etc.</p>		<p>Surveys/Questionnaires →</p> <p>Self-report measures →</p> <p>Interviews →</p> <p style="text-align: center;">Without scales, checklists, etc.</p>		
<p>Illuminative <i>Designed to measure whatever respondents choose – examines their categories on their terms</i></p>	<p>N/A</p>		<p>Written responses</p> <p>Oral responses</p>		

Table 2 – Examples of Various Types of Assessment Instruments

- **Pretest-Post Test: Water Unit** – <http://csf.concord.org/esf/curriculum>
Selected response, psychometric instrument. Administered as a pre- and post-test in a summative assessment project.
- **Pre-test/Post-test** – <http://csf.concord.org/esf/curriculum>
Constructed response, psychometric instrument. Administered as a pre- and post-test in a summative assessment project.
- **Chemistry Comprehension Exercise** – <http://www.umich.edu/~csie/undergraduate/jason/pretest>
Hybrid constructed and selected response, psychometric instrument. Could be used for formative or summative assessment.
- **Beginning Sexual Attitudes Survey for Elementary, Middle & High School Students** –
<http://abstinence.ssw.fsu.edu/traditional-survey.pdf>
Selected response, sociological instrument. Administered as a pre- and post-test in a summative assessment project.
- **Business: What It's All About?** – http://wps.prenhall.com/bp_reimers_finacct_1/0,,150885-,00.html and
http://wps.prenhall.com/bp_reimers_finacct_1/0,,150891-,00.html
Selected response, psychometric instrument. Pre- and post-tests separate instruments.
- **California State University: Information Competence Assessment Project** –
<http://www.csupomona.edu/~kkdunn/lcassess/phase2tasks.htm>
Example of an alternative assessment project; used interviews with and observations of participants.
- **Student Course Portfolios: Format and Guidelines** -
<http://www.chems.msu.edu/classes/s02/312/Portfolio.pdf>
Example of an alternative assessment project; guidelines to be used by students in completing a portfolio in chemical engineering and materials science. Uses as a summative assessment project.
 - See also **Class Portfolio Guidelines** for ENGR 350 -
<http://www.cecs.csulb.edu/~jewett/engr350/crqguide.html>
 - See also completed **Electronic Portfolios** -
<http://www.mehs.educ.state.ak.us/portfolios/portfolio.html>
- **Physics Lab Rubric** - <http://www.fayar.net/east/teacher.web/science/reif/Rubrics/labrubric.PDF>
Example of an alternative assessment project; rubric to be used in evaluating student performance on physics laboratory activities.
- **Engineering Student Attitude Assessment** - <http://www.engrng.pitt.edu/~outcomes/>
Selected response, sociological instrument. Used in a formative assessment project.
- **Chemical Engineering Department Undergraduate Assessment Program** -
http://www.mines.edu/fs_home/rmillier/measures.pdf
Rubric for a summative outcomes assessment project
- **An Intelligent Tutor for Classical Physics** - <http://www.press.umich.edu/jep/06-01/schulze.html>
Includes formative, self-assessment opportunities for students.