Chapter 8
Using Assessment Results to Inform Teaching Practice and Promote Lasting Learning

Linda Suskie

Introduction

While some may view systematic strategies to assess student learning as merely chores to satisfy quality assurance agencies and other external stakeholders, for faculty who want to foster lasting learning assessment is an indispensable tool that informs teaching practice and thereby promotes lasting learning. Suskie (2004c) frames this relationship by characterizing assessment as part of a continual four-step teaching-learning-assessment cycle.

The first step of this cycle is articulating expected learning outcomes. The teaching-learning-process is like taking a trip – one cannot plot out the route (curriculum and pedagogies) without knowing the destination (what students are to learn). Identifying expected student learning outcomes is thus the first step in the process. The more clearly the expected outcomes are articulated (i.e., articulating that one plans to visit San Francisco rather than simply the western United States), the easier it is to assess whether the outcome has been achieved (i.e., whether the destination has been reached).

The second step of the teaching-learning-assessment cycle is providing sufficient learning opportunities, through curricula and pedagogies, for students to achieve expected outcomes. Students will not learn how to make an effective oral presentation, for example, if they are not given sufficient opportunities to learn about the characteristics of effective oral presentations, to practice delivering oral presentations, and to receive constructive feedback on them.

The third step of the teaching-learning-assessment cycle is assessing how well students have achieved expected learning outcomes. If expected learning outcomes are clearly articulated and if students are given sufficient opportunity to achieve those outcomes, often this step is not particularly difficult – students' learning opportunities become assessment opportunities as well. Assignments in which students prepare and deliver oral presentations, for example, are not just opportunities for them to hone their oral presentation skills but also

L. Suskie
Middle States Commission on Higher Education, Philadelphia, PA, USA
e-mail: LSuskie@msche.org

G. Joughin (ed.), *Assessment, Learning and Judgement in Higher Education*,
DOI: 10.1007/978-1-4020-8905-3_8, © Springer Science+Business Media B.V. 2009

opportunities for faculty to assess how effectively students have developed those skills.

The final step of the teaching-learning-assessment cycle is using results to inform teaching-learning practice and thereby promote lasting learning. As Dochy has noted in Chapter 6, assessments can be viewed as tools to enhance the instructional process. Indeed, the best assessments have what Dochy, along with Linn and Dunbar (1991) and Messick (1989, 1994), have called *consequential validity* and what Moran and Malott (2004) have called *pedagogical validity* – assessment results are used as the basis for appropriate action and, specifically, to help "achieve the instructional objectives" (Moran & Malott, 2004, p. 137).

This chapter explores this conception of assessment as a means of promoting lasting learning through the consideration of three topics. First, teaching practices that have been shown through research to promote deep, lasting learning are reviewed. Next, some key underlying principles for using assessment results to inform teaching practices are discussed. Finally, practical suggestions for using assessment results to inform teaching practice and promote lasting learning are offered and explained through examples drawn from three very different kinds of assessment tools: rubrics (rating scales or scoring guides), multiple choice tests, and qualitative assessments such as reflective writing.

Teaching Practices that Promote Deep, Lasting Learning

Today's faculty are, in many ways, living in a golden age of education: their teaching practices can be informed by several decades of extensive research and publications documenting teaching practices that promote deep, lasting learning (e.g., Angelo, 1993; Association of American Colleges and Universities, 2002; Astin, 1993; Barr & Tagg, 1995; Chickering & Gamson, 1987, 1991; Ewell & Jones, 1996; Huba & Freed, 2000; Kuh, 2001; Kuh, Schuh, Whitt, & Associates, 1991; Light, 2001; McKeachie, 2002; Mentkowski & Associates, 2000; Palmer, 1998; Pascarella, 2001; Pascarella & Terenzini, 2005; Romer, & Education Commission of the States, 1996). Suskie (2004b) has aggregated this work into a list of 13 conditions under which students learn most effectively (p. 311):

1. Students understand course and program goals and the characteristics of excellent work.
2. They are academically challenged and given high but attainable expectations.
3. They spend more time actively involved in learning and less time listening to lectures.
4. They engage in multidimensional real-world tasks in which they explore, analyze, justify, evaluate, use other thinking skills, and arrive at multiple solutions. Such tasks may include realistic class assignments, field experiences, and service-learning opportunities.

5. The diversity of their learning styles is respected; they are given a variety of ways to learn and to demonstrate what they've learned.
6. They have positive interactions with faculty and work collaboratively with fellow students.
7. They spend significant time studying and practicing.
8. They receive prompt, concrete feedback on their work.
9. They have opportunities to revise their work.
10. They participate in co-curricular activities that build on what they are learning in the classroom.
11. They reflect on what and how they have learned and see coherence in their learning.
12. They have a synthesizing experience such as a capstone course, independent study, or research project.
13. Assessments focus on the most important course and program goals and are learning activities in their own right.

These principles elucidate two ways that assessment activities play a central role in promoting deep, lasting learning. First, as Dochy has noted in Chapter 6, providing a broad array of activities in which students learn and are assessed (Principle 5) promotes deep, lasting learning (Biggs, 2001; Entwistle, 2001). This rich array of evidence also provides a more complete and more meaningful picture of student learning, making assessment evidence more usable and useful in understanding and improving student learning. At its best, this broad array of learning/assessment activities is designed to incorporate three other principles for promoting lasting learning:

Principle 5: Students learn more effectively when the diversity of their learning styles is respected and they are given a variety of ways to learn and to demonstrate what they've learned. Suskie (2000) has noted that, because every assessment is inherently imperfect, any decisions related to student learning should be based on multiple sources of evidence. Furthermore, because every assessment favors some learning styles over others, students should have "a variety of ways to demonstrate what they've learned" (p. 8).

Principle 11: Students learn more effectively when they reflect on what and how they have learned and see coherence in their learning. In Chapter 4, Sadler has argued extensively on the educational value of students' practicing and developing skill in critical self-appraisal. Self-reflection fosters metacognition: learning how to learn by understanding how one learns (Suskie, 2004a).

Principle 12: Students learn more effectively when they have a synthesizing experience such as a capstone course, independent study, or research project. Such experiences give students opportunities to engage in synthesis – integrating what they have learned over their academic careers into a new whole (Henscheid, 2000).

The second way that assessment activities play a central role in promoting deep, lasting learning is through the time and energy students spend learning what they will be graded on, as Dochy has discussed extensively in Chapter 6. Indeed, Entwistle (2001) asserts, "It is not the teaching-learning environment

itself that determines approaches to studying, but rather what students believe to be required" (p. 16). Biggs (2001) concurs, noting, "[Students] see the assessment first, then learn accordingly, then at last see the outcomes we are trying to impart" (p. 66). The assessments that are used to grade students can thus be a powerful force influencing what and how students learn. This idea is captured in the final principle: "Assessments focus on the most important course and program goals and are learning activities in their own right." Four of the principles are keys to making this happen:

Principle 1: Students learn more effectively when they understand course and program goals and the characteristics of excellent work. As Sadler has discussed in Chapter 4, students who want to do well learn more effectively when they understand clearly why they have been given a particular assignment, what they are expected to learn from it, and how they will be graded on it. If students know, for example, that their papers will be graded in part on how effective the introduction is, some will try their best to write an effective introduction.

Principle 2: Students learn more effectively when they are academically challenged and given high but attainable expectations. While there is a limit to what students can achieve in a given amount of time (even the best writing faculty, for example, cannot in a few weeks enable first-year students to write at the level expected of doctoral dissertations), many students respond remarkably well to high standards, provided that they are given a clear roadmap on how to achieve them (De Sousa, 2005). First-year students, for example, may be able to write quite competent library research papers if the research and writing process is broken down into relatively small, manageable steps (identifying the research topic, finding appropriate library resources, reading and analyzing them, outlining the paper, etc.) and students are guided through each step.

Principle 7: Students learn more effectively when they spend significant time studying and practicing. While this is an obvious truism, there is evidence that students in some parts of the world spend relatively little time on out-of-class studies. Those who attended college a generation or two ago may recall the adage to study two hours outside of class for every hour spent in class. For a full-time student spending 15 hours in class, this would mean spending about 30 hours on out-of-class studies. But according to the National Survey of Student Engagement (2006), about 45% of today's American college freshmen and seniors report spending less than 11 hours a week preparing for class (studying, reading, writing, doing homework or lab work, analyzing data, rehearsing, and other academic activities). Only about ten percent spend more than 25 hours a week preparing for class. While these results reflect part-time as well as full-time students, it is nonetheless clear that, at least in the United States, much more could be done to have students take responsibility for their learning.

Principle 8: Students learn more effectively when they receive prompt, concrete feedback on their work. As Dochy has discussed in Chapter 6, we all benefit from constructive feedback on our work, and students are no different (Ovando, 1992). The key is to get feedback to students quickly enough that they can use the feedback to improve their learning. Faculty know all too well that final

exams graded after classes end are often not retrieved by students and, if they are, checked only for the final grade. The problem is that providing constructive, timely feedback can take a great deal of time. Here are some practical suggestions on ways to minimize that time:

- Return ungraded any assignments that show little or no effort, with a request that the paper be redone and resubmitted within 24 hours. As Walvoord and Anderson (1998) point out, if students make no effort to do an assignment well, why should the professor make any effort to offer feedback?
- When appropriate (see Sadler, Chapter 4 of this book) use rubrics (rating scales or scoring guides) to evaluate student work. They speed up the process because much feedback can be provided simply by checking or circling appropriate boxes rather than writing comments.
- Use Haswell's (1983) *minimal marking method*: rather than correct grammatical errors, simply place a check in the margin next to the error, and require the student to identify and correct errors in that line.
- Provide less feedback on minor assignments. Some short homework assignments, for example, can be marked simply with a plus symbol for outstanding work, a checkmark for adequate work, and a minus symbol for minimal effort.

Using Assessment Results to Promote Lasting Learning: Two Underlying Principles

Assessment results can provide a wealth of information to help faculty understand how effective their teaching is and how it might be improved, provided that two principles are followed when the assessments are planned.

Articulate the Decisions that Assessment Will Inform

MacGregor, Tinto, and Lindblad (2001) note, "If you're not clear on the goals of an assessment and the audiences to which that assessment will be directed, it's hard to do the assessment well. So your first task is to ask yourself why, with whom, and for what purpose you are assessing…" (p. 48). Assessment results can inform decisions in at least five areas:

Learning outcomes. Assessment results might help faculty decide, for example, if their statements of expected learning outcomes are sufficiently clear and focused or if they have too many intended learning outcomes to cover in the allotted instructional time.

Curriculum. Assessment results might help faculty decide, for example, if classes or modules taught by several faculty have sufficient uniformity across sections or whether a service-learning component is achieving its goal.

Pedagogy. Assessment results might help faculty decide, for example, whether online instruction is as effective as traditional classroom-based instruction or whether collaborative learning is more effective than traditional lectures.

Assessment. Assessment results can, of course, help faculty decide how useful their assessment strategies have been and what changes are needed to improve their effectiveness.

Resource allocations. Assessment results can provide a powerful argument for resource allocations. Disappointing evidence of student writing skills, for example, can lead to fact-based arguments for, say, more writing tutors or more online writing software. Poor student performance on a technology examination required for licensure in a profession can be a compelling argument for upgrading technologies in laboratories.

Understanding and articulating the decisions that a particular assessment will inform helps to ensure that the assessment results will indeed help enlighten those decisions. Suppose, for example, that a professor is assessing student learning in a biology laboratory. If the results are to inform decisions about the laboratory's learning outcomes, for example, the assessments will need to be designed to assess each expected outcome. If the results are to inform decisions about curriculum, the assessments will need to be designed to assess each aspect of the curriculum. And if the results are to inform decisions about pedagogy, the assessments may need to be designed to provide comparable information on different pedagogical approaches.

Develop Assessment Strategies that Will Provide Appropriate Frames of Reference to Inform Those Decisions

Suskie (2007) has observed that assessment results considered in isolation are meaningless – they have significance only if they are compared against some kind of benchmark or frame of reference. Suskie has identified nine such frames of reference, four of which are especially relevant to most faculty.

Under the *strengths and weaknesses* frame of reference, faculty compare the sub-scores of an assessment to identify students' relative strengths and weaknesses. An assessment of writing skills, for example, might determine that students are relatively good at writing a strong introduction but relatively weak at supporting arguments with appropriate evidence. This frame of reference is often of great interest and great value to faculty, as it tells them what their students are learning well and what areas need more or different attention. In order to use this frame of reference, the assessment must be designed to generate comparable results on aspects of the trait being assessed, such as effective writing. This is generally accomplished by using a rubric or rating scale that lists the aspects being evaluated. Some published tests and surveys also generate sub-scores that can be compared with one another. In contrast, a holistic assessment, generating only one score per student without subscores, cannot provide this kind of information.

Under the *improvement* frame of reference, faculty compare student assessment results at the beginning and end of a class, module, course, or program. Faculty might, for example, give students the final examination on the first day of class and compare those scores against those on the same examination given at the end of instruction. Or faculty might give students writing assignments in the first and last weeks that are evaluated using the same criteria.

Such a value-added approach is intrinsically appealing, as it appears to convey how much students' learning has improved as a result of faculty teaching. But this approach has a number of shortcomings. One is that students must be motivated to do their best on the entry-point assessment. This can be a challenge because grading, generally a strong motivator, is often inappropriate at this point: Is it fair to grade students when they have not yet had an opportunity to learn anything?

Another challenge is that both entry- and exit-point assessment information must be collected for this approach to be meaningful. This is not possible in situations in which sizable numbers of students either transfer into a class, module, course, or program after it has begun or drop out of it before it is completed. In these situations, the students who persist from beginning to end may not be a representative sample of all students who are subject to instruction.

Yet another challenge with this value-added approach is that gain scores, the difference between entry- and exit-point assessment results, are notoriously unreliable. As noted by Banta and Pike (2007) and Pike (2006), the measurement error of gain scores is essentially double that of each assessment alone. This sizable measurement error can mask meaningful gains.

But perhaps the major concern about the improvement or value-added frame of reference is that it is often confused with the pre-post experimental design (Campbell & Stanley, 1963) used in the social sciences. In pre-post experimental designs, subjects are randomly assigned to control and experimental treatments; this allows the research to separate the impact of the treatment from extraneous factors. In higher education, however, faculty cannot randomly assign students to institutions or programs, so if faculty find significant growth they cannot conclude that it is due to the learning experience or to extraneous factors. If a student's oral communication skills improve, for example, faculty cannot be certain that the improvement is due to work in class or to, say, concurrent participation in a club or a part-time job in which the student uses and improves oral communication skills.

Under the *historical trends* frame of reference, faculty compare student assessment results against those of prior cohorts of students. This frame of reference is of particular interest to faculty who want to know if their efforts to improve their curricula and pedagogies are yielding desired improvements in student learning. This frame of reference can only be used, of course, if identical or parallel assessments can be utilized with successive cohorts of

students. This is not always possible – sometimes curricula must change in order to meet employer and societal demands, so assessments used a few years ago may no longer be appropriate or relevant today. Another challenge with this approach is that, as with the improvement frame of reference discussed above, this is not an experimental design with random assignments to cohorts. As a result, faculty cannot be certain that changes in student learning are due to changes in curricula or pedagogies or due to changes in the students themselves. Faculty teaching at an institution that has increased its admission standards, for example, cannot be certain that the growth they see in student learning is due to changes in curricula and pedagogies or is simply because today's students are more talented, motivated, or prepared to learn.

Under the *standards* frame of reference, faculty compare assessment results against an established standard, set either by the faculty or by a regional or national agency or organization. Faculty might decide for example, that students must answer at least 70% of test questions correctly in order to pass an examination, or a nursing licensure agency might state that nursing students must earn a particular score on a licensure examination in order to be licensed to practice. This frame of reference is of interest to faculty who want or need to ensure that students are meeting particular standards. Many colleges and universities, for example, want to ensure that all students graduate with a particular level of writing skill.

The challenge with this approach is, not surprisingly, setting an appropriate standard. If the standard has been set by an agency or organization, the work is done, of course, but setting a defensible, valid local standard can be very difficult and time-consuming. While faculty have been doing this for generations (setting a standard, for example, that students must answer at least 65% of questions correctly in order to pass a final examination), in reality these standards are often set arbitrarily and without clear justification. Livingston and Zieky (1982) offer a variety of techniques for establishing defensible standards.

Practical Suggestions for Summarizing, Interpreting, and Using Assessment Results To Whom It May Concern: Promote Deep, Lasting Learning

In order to use assessment results to inform teaching practice and thereby improve student learning, they must be summarized in a way that busy faculty can quickly and easily understand. They must then be interpreted in appropriate ways so they may be used to inform teaching practice and thereby promote deep, lasting learning. Summaries, analyses, and interpretations should aim to answer two fundamental questions: (1) What have we learned about our students' learning? and (2) What are we going to do about what we

have learned? The key is to ensure that the steps that are taken build upon practices that promote deep, lasting learning.

What follow are practical suggestions for summarizing, interpreting, and using assessment results to answer these questions for three different assessment practices: the use of rubrics (rating scales or scoring guides), multiple choice tests, and reflective writing exercises.

Rubrics

As Sadler has discussed in Chapter 4, rubrics are a list of the criteria used to evaluate student work (papers, projects, performances, portfolios, and the like) accompanied by a rating scale. Using rubrics can be a good pedagogical practice for several reasons:

- Creating a rubric *before* the corresponding assignment is developed, rather than vice versa, helps to ensure that the assignment will address what the professor wants students to learn.
- Giving students the rubric along with the assignment is an excellent way to help them understand the purpose of the assignment and how it will be evaluated.
- Using a rubric to grade student work ensures consistency and fairness.
- Returning the marked rubric to students with their graded assignment gives them valuable feedback on their strengths and weaknesses and helps them understand the basis of their grade.

In order to understand rubric results and use them to inform teaching practice, it is helpful to tally students' individual ratings in a simple chart. Table 8.1 provides a hypothetical example for the results of a rubric used to evaluate the portfolios of 30 students studying journalism.

It is somewhat difficult to understand what Table 8.1 is saying. While it seems obvious that students performed best on the fourth criterion ("understanding professional ethical principles and working ethically"), their other relative strengths and weaknesses are not as readily apparent. Table 8.1 would be more useful if the results were somehow sorted. Let us suppose that, in this hypothetical example, the faculty's goal is that all students earn at least "very good" on all criteria. Table 8.2 sorts the results based on the number of students who score either "excellent" or "very good" on each criterion. Table 8.2 also converts the raw numbers into percentages, because this allows faculty to compare student cohorts of different sizes. The percentages are rounded to the nearest whole percentage, to reduce the volume of information to be digested and to keep the reader from focusing on trivial differences.

Now the results jump out at the reader. It is immediately apparent that students not only did relatively well on the fourth criterion but also on the first, third and fifth. It is equally apparent that students' weakest area (of those evaluated here) is the tenth criterion ("applying basic numerical and statistical

Table 8.1 An example of tallied results of a rubric used to evaluate the portfolios of 30 students studying journalism

	The student:	Excellent	Very Good	Adequate	Inadequate
1.	Understands and applies the principles and laws of freedom of speech and press.	15	14	1	0
2.	Understands the history and role of professionals and institutions in shaping communications.	18	8	4	0
3.	Understands the diversity of groups in a global society in relationship to communications.	12	17	1	0
4.	Understands professional ethical principles and works ethically in pursuit of truth, accuracy, fairness, and diversity.	27	3	0	0
5.	Understands concepts and applies theories in the use and presentation of images and information.	12	17	1	0
6.	Thinks critically, creatively, and independently.	2	20	5	3
7.	Conducts research and evaluates information by methods appropriate to the communications profession(s) studied.	6	21	3	0
8.	Writes correctly and clearly in forms and styles appropriate for the communications profession(s) studied and the audiences and purposes they serve.	9	14	6	1
9.	Critically evaluates own work and that of others for accuracy and fairness, clarity, appropriate style and grammatical correctness.	6	19	3	2
10.	Applies basic numerical and statistical concepts.	10	8	9	3
11.	Applies tools and technologies appropriate for the communications profession(s) studied.	8	19	3	0

concepts"). Another area of relative weakness is the sixth criterion ("thinking critically, creatively, and independently"). Table 8.2 thus provides a clear roadmap for faculty reflection on students' relative strengths and weaknesses.

Professors might begin this reflection by first examining how the curriculum addresses the application of basic numerical and statistical concepts by reviewing syllabi to identify the courses or modules in which this skill is addressed. The faculty might then discuss how well the identified courses or modules follow the

Table 8.2 An improved version of Table 8.1

	The student:	Excellent + Very good	Excellent	Very good	Adequate	Inadequate
3.	Understands the diversity of groups in a global society in relationship to communications.	97%	40%	57%	3%	0%
5.	Understands concepts and applies theories in the use and presentation of images and information.	97%	40%	57%	3%	0%
11.	Applies tools and technologies appropriate for the communications profession(s) studied.	90%	27%	63%	10%	0%
7.	Conducts research and evaluates information by methods appropriate to the communications profession(s) studied.	90%	20%	70%	10%	0%
2.	Understands the history and role of professionals and institutions in shaping communications.	87%	60%	27%	13%	0%
9.	Critically evaluates own work and that of others for accuracy and fairness, clarity, appropriate style and grammatical correctness.	83%	20%	63%	10%	7%
8.	Writes correctly and clearly in forms and styles appropriate for the communications profession(s) studied and the audiences and purposes they serve.	77%	30%	47%	20%	3%
6.	Thinks critically, creatively, and independently.	73%	7%	67%	17%	10%
10.	Applies basic numerical and statistical concepts.	60%	33%	27%	30%	10%

thirteen principles for promoting deep, lasting learning articulated in this chapter. They might, for example, ask themselves:

- Are we giving enough time and attention in these classes to applying basic numerical and statistical concepts? Are we giving students enough classwork and assignments on this skill?
- Do students spend enough time actively applying numerical and statistical concepts?
- Are the assignments in which they apply numerical and statistical concepts real world problems, the kinds that may have more than one "correct" answer?
- Would students benefit from working with fellow students on these assignments rather than alone?
- Do we give students sufficient feedback on their work in applying numerical and statistical concepts? Do we give them sufficient opportunities to correct or revise their work?

Discussion of these and similar questions will doubtless lead to ideas about ways to strengthen students' skills in applying numerical and statistical concepts. Faculty might decide, for example, to incorporate the application of numerical and statistical concepts into more courses or modules, to give students more practice through additional homework assignments, and to give students more collaborative projects in which they must interpret real world data with their fellow students.

Multiple Choice Tests

The validity and usefulness of multiple choice tests is greatly enhanced if they are developed with the aid of a test blueprint or outline of the knowledge and skills being tested. Table 8.3 is an example of a simple test blueprint.

In this example, the six listed objectives represent the professor's key objectives for this course or module, and the third, fourth, and sixth objectives are considered the most important objectives. This blueprint is thus powerful evidence of the content validity of the examination and a good framework for summarizing the examination results, as shown in Table 8.4. Again, the results in Table 4 have been sorted from highest to lowest to help readers grasp the results more quickly and easily.

Table 8.3 Example of a test blueprint for a statistics examination

1 item	Determine the value of t needed to find a confidence interval of a given size.
1 item	Understand the effect of p on the standard error of a proportion.
6 items	Choose the appropriate statistical analysis for a given research problem.
4 items	Decide on the appropriate null and alternative hypotheses for a given research problem and state them correctly.
2 items	Identify the critical value(s) for a given statistical test.
4 items	Choose the appropriate standard error formula for a given research problem.

Table 8.4 Results of a statistics examination, matched to the test blueprint

Percentage of students answering correctly	Learning Objective
95%	Determine the value of t needed to find a confidence interval of a given size.
88%	Understand the effect of p on the standard error of a proportion.
85%	Decide on the appropriate null and alternative hypotheses for a given research problem and state them correctly.
79%	Identify the critical value(s) for a given statistical test.
62%	Choose the appropriate standard error formula for a given research problem.
55%	Choose the appropriate statistical analysis for a given research problem

Table 8.4 makes clear that students overall did quite well in determining the value of t needed to find a confidence interval, but a relatively high proportion were unable to choose the appropriate statistical analysis for a given research problem. Table 8.4 provides another clear roadmap for faculty reflection on students' relative strengths and weaknesses. The professor might consider how to address the weakest area – choosing appropriate statistical analyses – by again reflecting on the practices that promote deep, lasting learning discussed earlier in this chapter. The professor might, for example, decide to address this skill by revising or expanding lectures on the topic, giving students additional homework on the topic, and having students work collaboratively on problems in this area.

Another useful way to review the results of multiple choice tests is to calculate what testing professionals (e.g., Gronlund, 2005; Haladyna, 2004; Kubiszyn & Borich, 2002) call the discrimination of each item (Suskie, 2004d). This metric, a measure of the internal reliability or internal consistency of the test, is predicated on the assumption that students who do relatively well on a test overall will be more likely to get a particular item correct than those who do relatively poorly. Table 8.5 provides a hypothetical example of discrimination results for a 5-item quiz. In this example, responses of the ten students with the highest overall scores on this quiz are compared against those of the ten students with the lowest overall quiz scores.

Table 8.5 Discrimination results for a short quiz taken by 30 students

Item number	Number of "Top 10" Students answering correctly	Number of "Bottom 10" Students answering correctly	Difference (Discrimination)
1	10	0	10
2	8	6	2
3	5	5	0
4	10	10	0
5	4	8	−4

These five items have widely varying levels of discrimination:

- Item 1 has the best possible discrimination – all the top students answered it correctly, while none of the bottom students did. This is truly an item that "separates the wheat from the chaff," discriminating students who have truly mastered class objectives from those who have not.
- Item 2 is an example of an item with good discrimination, though not as strong as Item 1, simply because it is easier. Items that fifty per cent of students get wrong have the greatest potential for discrimination; easier items have lower potential for discrimination.
- Item 3 has no discrimination – equal numbers of top and bottom students answered it correctly. While an item with no discrimination is not inherently a poor item, it would be worth a closer look to see why a number of top students struggled with it while some bottom students did not. Asking the top students why they got this question wrong would probably give the professor ideas on ways to revise the question for future administrations.
- Item 4 also has no discrimination, but this is simply because everyone answered it correctly. As already noted, easy items cannot discriminate well between top and bottom students, and items that are so easy that everyone answers them correctly will, of course, not discriminate at all.
- Item 5 discriminates negatively – students in the top group were more likely to answer it incorrectly than students in the bottom group. It is very likely that students in the top group misinterpreted either the question or one or more of its options, probably reading more into the item than the professor intended. This is an item that performed so poorly that it should be removed from the scores of these students and revised before it is used again. As with Item 3, the top students who got this question wrong would doubtless be able to give the professor suggestions on how to revise the item to minimize future misinterpretations.

Reflective Writing

Reflective writing is a learning strategy in which students reflect and write on what and how they have learned. Students engaged in reflective writing typically reflect and write on "the larger context, the meaning, and the implications of an experience or action" (Branch & Paranjape, 2002, p. 1185) and "pull together a broad range of previous thinking or knowledge in order to make greater sense of it for another purpose that may transcend the previous bounds of personal knowledge or thought" (Moon, 2001, p. 5). Reflective writing thus helps students develop a number of skills (Costa & Kallick, 2000), including skill in synthesis—pulling together what they have learned in order to see the big picture – and metacognition – understanding how one learns.

Reflective writing can also be a valuable assessment strategy. Costa and Kallick (2002) note that reflective writing provides an opportunity for

"documenting learning and providing a rich base of shared knowledge" (p. 60), while the Conference on College Composition and Communication notes that "reflection by the writer on her or his own writing processes and performances holds particular promise as a way of generating knowledge about writing" (2006). Reflective writing may be especially valuable for assessing ineffable outcomes such as attitudes, values, and habits of mind. An intended student learning outcome to "be open to diverse viewpoints" would be difficult to assess through a traditional multiple choice test or essay assignment, because students would be tempted to provide what they perceive to be the "correct" answer rather than accurate information on their true beliefs and views.

Because reflective writing seeks to elicit honest answers rather than "best" responses, reflective writing assignments may be assessed and the results used differently than other assessment strategies. While the structure of a student's reflective writing response can be evaluated using a rubric, the thoughts and ideas expressed may be so wide-ranging that qualitative rather than quantitative assessment strategies may be more appropriate.

Qualitative assessment techniques are drawn from qualitative research approaches, which Marshall and Rossman (2006) describe as "naturalistic," "fundamentally interpretive," relying on "complex reasoning that moves dialectically between deduction and induction," and drawing on "multiple methods of inquiry" (p. 2). Qualitative assessment results may thus be summarized differently than quantitative results such as those from rubrics and multiple choice tests, which typically yield ratings or scores that can be summarized using descriptive and inferential statistics. Qualitative research techniques aim for naturalistic interpretations rather than, say, an average score.

Qualitative assessment techniques typically include sorting the results into categories (e.g., Patton, 2002). Tables 8.6 and 8.7 provide an example of a summary of qualitative assessment results from a day-long workshop on the assessment of student learning, conducted by the author. The workshop addressed four topics: principles of good practice for assessment, promoting an institutional culture of assessment, the articulation of learning outcomes, and assessment strategies including rubrics. At the end of the day, participants were asked two questions, adapted from the *minute paper* suggested by Angelo

Table 8.6 Responses to "What was the most useful or meaningful thing you learned today?" by Participants at a one-day workshop on assessing student learning

Percent of respondents (%)	Category of response
40	Assessment strategies (e.g., rubrics)
20	Culture of assessment
16	Principles of good practice
10	Articulating learning outcomes
13	Miscellaneous

Table 8.7 Responses to "What question remains uppermost on your mind as we end this workshop?" by participants at a one-day workshop on assessing student learning

Percent of Respondents (%)	Category of Response
27	Culture of assessment
13	Organizing assessment across an institution
43	Unique questions on other topics
16	No response

and Cross (1993): "What was the most useful or meaningful thing you learned today?" and "What one question is uppermost on your mind as we end this workshop?"

For the first question, "What was the most useful or meaningful thing you learned today?" (Table 8.6), the author sorted comments into five fairly obvious categories: the four topics of the workshop plus a "miscellaneous" category. For the second question, "What one question is uppermost on your mind as we end this workshop?" (Table 8.7), potential categories were not as readily evident. After reviewing the responses, the author settled on the categories shown in Table 8.7, then sorted the comments into the identified categories. Qualitative analysis software is available to assist with this sorting – such programs search responses for particular keywords provided by the professor.

The analysis of qualitative assessment results – identifying potential categories for results and then deciding the category into which a particular response is placed – is, of course, inherently subjective. In the workshop example described here, the question, "Would it be helpful to establish an assessment steering committee composed of faculty?" might be placed into the "culture of assessment" category by one person and into the "organizing assessment" category by another. But while qualitative assessment is a subjective process, open to inconsistencies in categorizations, it is important to note that *any* kind of assessment of student learning has an element of subjectivity, as the questions that faculty choose to ask of students and the criteria used to evaluate student work are a matter of professional judgment that is inherently subjective, however well-informed. Inconsistencies in categorizing results can be minimized by having two readers perform independent categorizations, then reviewing and reconciling differences, perhaps with the introduction of a third reader for areas of disagreement.

Qualitative assessment results can be extraordinarily valuable in helping faculty understand and improve their teaching practices and thereby improve student learning. The "Minute Paper" responses to this workshop (Tables 8.6 and 8.7) provided a number of useful insights to the author:

- The portion of the workshop addressing assessment strategies was clearly very successful in conveying useful, meaningful ideas; the author could take satisfaction in this and leave it as is in future workshops.

- The portion of the workshop addressing learning outcomes was not especially successful; while there were few if any questions about this topic, few participants cited it as especially useful or meaningful. (A *background knowledge probe* (Angelo & Cross, 1993) would have revealed that most participants arrived at the workshop with a good working knowledge of this topic.) The author used this information to modify her workshop curriculum to limit coverage of this topic to a shorter review.
- Roughly one in eight participants had questions about organizing assessment activities across their institution, a topic not addressed in the workshop. The author used this information to modify her workshop curriculum to incorporate this topic. (Reducing time spent on learning outcomes allowed her to do this.)

The portion of the workshop addressing promoting a culture of assessment was clearly the most problematic. While a fifth of all respondents found it useful or meaningful, two-fifths had questions about this topic when the workshop concluded. Upon reflection, the author realized that she placed this topic at the end of the workshop curriculum, addressing it at the end of the day when she was rushed and participants were tired. She modified her curriculum to move this topic to the beginning of the workshop and spend more time on it.

As a result of this reflective writing assignment and the subsequent changes made in curriculum and pedagogy, participant learning increased significantly in subsequent workshops, as evidenced by the increased proportion of comments citing organizing assessment activities and promoting a culture of assessment as the most useful or meaningful things learned and the smaller proportion of participants with questions about these two areas.

Conclusion

Why do faculty assess student learning? One longstanding reason, of course, is to form a basis for assigning grades to students. Another recently emerging reason is to demonstrate to various constituents – government agencies, quality assurance agencies, taxpayers, employers, and students and their families – that colleges and universities are indeed providing students with the quality education they promise. But the most compelling reason for many faculty to engage in assessing student learning is for the opportunity it provides to improve teaching practices and thereby foster deep, lasting learning. This chapter has described a number of ways that assessment activities can accomplish these ends:

- Provide a broad array of learning and assessment activities.
- Design assessment activities (e.g., assignments, tests) so that they address key learning outcomes.
- Help students understand course and program expectations and the characteristics of excellent work.

- Challenge students by giving them high but attainable expectations.
- Require students to spend significant time studying and practicing.
- Give students prompt, concrete feedback on their work.
- Articulate the decisions that assessment results are to inform.
- Design assessments so that they will provide appropriate frames of reference to inform those decisions.
- Use test blueprints to plan multiple choice tests and rubrics to plan other assignments and tests.
- Summarize assessment results into simple tables, perhaps with results sorted so that the best and most disappointing results can be quickly identified.
- Use the results of rubrics and multiple choice tests to identify students' relative strengths and weaknesses and ways that the assessments themselves might be improved.
- Use the results of qualitative assessments to identify areas in which students are confused, dissatisfied with their learning, or fail to demonstrate attainment of key learning outcomes.
- Use recent research on strategies that promote deep, lasting learning, along with feedback from students, to plan how to address assessment results that are disappointing and thereby improve teaching practice and promote lasting learning.

Faculty who have a passion for teaching are always looking for ways to improve their practice and foster lasting student learning. Once they understand the nature and use of assessment, they quickly come to realize that assessment is one of the best tools in their teaching toolbox for achieving these ends.

References

Angelo, T. A. (1993, April). A "teacher's dozen": Fourteen general, research-based principles for improving higher learning in our classrooms. *AAHE Bulletin, 45*(8), 3–7, 13.

Angelo, T. A., & Cross, K. P. (1993). *Classroom assessment techniques: A handbook for college teachers* (2nd ed.). San Francisco: Jossey-Bass.

Association of American Colleges and Universities. (2002). *Greater expectations: A New vision for learning as a nation goes to college*. Washington, DC: Author.

Astin, A. W. (1993). *What matters in college: Four critical years revisited*. San Francisco: Jossey-Bass.

Banta, T. W., & Pike, G. R. (2007). Revisiting the blind alley of value added. *Assessment Update, 19*(1), 1–2, 14–15.

Barr, R. B., & Tagg, J. (1995). From teaching to learning: A new paradigm for undergraduate education. *Change, 27*(6), 12–25.

Biggs, J. (2001). Assessing for quality in learning. In L. Suskie (Ed.), *Assessment to promote deep learning: Insight from AAHE's 2000 and 1999 assessment conferences* (pp. 65–68).

Branch, W. T., & Paranjape, A. (2002). Feedback and reflection: Teaching methods for clinical settings. *Academic Medicine, 77*, 1185–1188.

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Chicago: Rand McNally.

Chickering, A. W., & Gamson, Z. (1987). Seven principles for good practice in undergraduate education. *AAHE Bulletin, 39*(7), 5–10.

Chickering, A. W., & Gamson, Z. (1991). *Applying the seven principles for good practice in undergraduate education*. New directions for teaching and learning, No. 47. San Francisco: Jossey-Bass.

Conference on College Composition and Communication. (2006). Writing assessment: A position statement. Urbana, IL: National Council of Teachers of English. Retrieved September 4, 2007, from http://www.ncte.org/cccc/resources/123784.htm

Costa, A., & Kallick B. (2000). Getting into the habit of reflection. *Educational Leadership, 57*(7), 60–62.

De Sousa, D. J. (2005). *Promoting student success: What advisors can do* (Occasional Paper No. 11). Bloomington, Indiana: Indiana University Center for Postsecondary Research.

Entwistle, N. (2001). Promoting deep learning through teaching and assessment. In L. Suskie (Ed.), *Assessment to promote deep learning: Insight from AAHE's 2000 and 1999 assessment conferences* (pp. 9–20).

Ewell, P. T., & Jones, D. P. (1996). *Indicators of "good practice" in undergraduate education: A handbook for development and implementation*. Boulder, CO: National Center for Higher Education Management Systems.

Gronlund, N. E. (2005). *Assessment of student achievement* (8th ed.) Boston: Allyn & Bacon.

Haladyna, T. M. (2004). *Developing and validating multiple choice items*. Boston, MA: Allyn & Bacon.

Haswell, R. (1983). Minimal marking. *College English, 45*(6), 600–604.

Henscheid, J. M. (2000). Professing the disciplines: An analysis of senior seminars and capstone courses (Monograph No. 30). Columbia, South Carolina: National Resource Center for the First-Year Experience and Students in Transition.

Huba, M. E., & Freed, J. E. (2000). *Learner-centered assessment on college campuses: Shifting the focus from teaching to learning* (pp. 32–64). Needham Heights, MA: Allyn & Bacon.

Kubiszyn, T., & Borich, G. D. (2002). *Educational testing and measurement: Classroom application and management* (7th ed.) San Francisco: Jossey-Bass.

Kuh, G. (2001). Assessing what really matters to student learning: Inside the National Survey of Student Engagement. *Change, 33*(3), 10–17, 66.

Kuh, G. D., Schuh, J. H., Whitt, E. J., & Associates. (1991). *Involving colleges: Successful approaches to fostering student learning and development outside the classroom*. San Francisco: Jossey-Bass.

Light, R. (2001). *Making the most of college: Students speak their minds*. Cambridge, MA: Harvard University Press.

Linn, R., & Dunbar, S. B. (1991). *Complex, performance-based assessment: Expectations and validation criteria*. Los Angeles: University of California Center for Research on Evaluation, Standards, and Student Testing.

Livingston, S. A., & Zieky, M. J. (1982). *Passing scores: A manual for setting standards on performance on educational and occupational tests*. Princeton: Educational Testing Service.

MacGregor, J., Tinto, V., & Lindblad, J. H. (2001). Assessment of innovative efforts: Lessons from the learning community movement. In L. Suskie (Ed.), *Assessment to promote deep learning: Insight from AAHE's 2000 and 1999 assessment conferences* (pp. 41–48).

McKeachie, W. J. (2002). *Teaching tips: Strategies, research, and theory for colleges and university teachers* (11th ed.). Boston: Houghton Mifflin.

Marshall, C., & Rossman, G. B. (2006). *Designing qualitative research* (4th ed.). Thousand Oaks, CA: Sage.

Mentkowski, M., & Associates. (2000). *Learning that lasts: Integrating learning, development, and performance in college and beyond*. San Francisco, CA: Jossey-Bass.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement*. New York: Macmillan.

Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher, 23*(2), 13–23.

Moon, J. (2001). *Reflection in learning and professional development: Theory and practice*. London: Routledge.
Moran, D. J., & Malott, R. W. (Eds.). (2004). *Evidence-based educational methods*. San Diego: Elsevier Academic Press.
National Survey of Student Engagement. (2006). *Engaged learning: Fostering success for all students: Annual report 2006*. Bloomington, IN: Author.
Ovando, M. N. (1992). Constructive feedback: A key to successful teaching and learning. Austin, TX: University of Texas at Austin, College of Education, Department of Education Administration. ERIC Document Reproductive Services No. ED 404 291.
Palmer, P. J. (1998). *The courage to teach: Exploring the inner landscape of a teacher's life*. San Francisco: Jossey-Bass.
Pascarella, E. T. (2001). Identifying excellence in undergraduate education: Are we even close? *Change, 33*(3), 19–23.
Pascarella, E. T., & Terenzini, P.T. (2005). *How college affects students: A third decade of research*. San Francisco: Jossey-Bass.
Patton, M. Q. (2002). *Qualitative research & evaluation methods* (3rd ed.). Thousand Oaks, CA: Sage.
Pike, G. R. (2006). Assessment measures: Value-added models and the Collegiate Learning Assessment. *Assessment Update, 18*(4), 5–7.
Romer, R., & Education Commission of the States. (1996, April). What research says about improving undergraduate education. *AAHE Bulletin, 48*(8), 5–8.
Suskie, L. (2000, May). Fair assessment practices: Giving students equitable opportunities to demonstrate learning. *AAHE Bulletin, 52*(9), 7–9.
Suskie, L. (2004a). Encouraging student reflection. In *Assessing student learning: A common sense guide* (pp. 168–184). San Francisco: Jossey-Bass Anker Series.
Suskie, L. (2004b). Using assessment findings effectively and appropriately. In *Assessing student learning: A common sense guide* (pp. 300–317). San Francisco: Jossey-Bass Anker Series.
Suskie, L. (2004c). What is assessment? Why assess? In *Assessing student learning: A common sense guide* (pp. 3–17). San Francisco: Jossey-Bass Anker Series.
Suskie, L. (2004d). Writing a traditional objective test. In *Assessing student learning: A common sense guide* (pp. 200–221). San Francisco: Jossey-Bass Anker Series.
Suskie, L. (2007). Answering the complex question of "How good is good enough?" *Assessment Update*, *19*(4), 1–2, 14–15.
Walvoord, B., & Anderson, V. J. (1998). *Effective grading: A tool for learning and assessment*. San Francisco: Jossey-Bass.