Assessment Clear and Simple

A Practical Guide for Institutions, Departments, and General Education

Second Edition

Barbara E. Walvoord

Foreword by Trudy W. Banta

JOSSEY-BASS
A Wiley Imprint
www.josseybass.com

Chapter 1

For Everyone

The Basics of Assessment

YOU PROBABLY ARE reading this book because you are an administrator, department chair, assessment director, general education committee member, or faculty member involved in assessment. I wrote this book after serving in several of those administrative and faculty roles myself and serving as a consultant on assessment for more than 350 institutions, public and private, large and small, traditional and nontraditional. I have written this book for all those people and their institutions.

The Purpose of This Book

This book provides a short, clear, no-nonsense guide to assessment. The book examines how assessment can serve departmental and institutional goals—not merely external mandates—and how assessment can be conducted effectively and efficiently with ordinary peoples' available time, expertise, and resources. This book aims to make assessment simple, cost efficient, and useful for student learning, while meeting the assessment requirements of accreditation agencies, legislatures, review boards, and others.

Relation to Other Resources

I have emphasized brevity and practicality. Other types of resources for assessment include collections of case studies such as Banta, Jones, and Black (2009) and longer books such as Suskie (2009), a very thorough guide at more than 300 pages. My and Anderson's *Effective Grading* (2010) focuses on classroom assessment, including establishing goals, designing assignments, encouraging student motivation, designing the course, communicating with students about their work, and saving time in the grading process. It forms a kind of Venn diagram with this book, because its final section discusses how to use student classroom work, as well as other measures, for assessment in departments or general education programs.

The Organization of This Book

This book is organized in the following way:

- This chapter, which everyone should read. It defines assessment, answers common concerns, and lays the groundwork for each of the following chapters.
- Chapter Two, for institution-wide leaders and planners: assessment directors and committees, provosts, deans, and anyone who wants to see the "big picture" for the institution.
- Chapter Three, for department members and chairs.
- Chapter Four, for general education leaders and faculty.

Themes of the Book

The following themes recur throughout this book:

- Assessment is a natural, scholarly act that can bring important benefits.
- Assessment is composed of three steps: goals, information, action.
- The end of assessment is action.
- Assessment involves communicating across cultures, within and outside the institution.
- You need not only individual data collection, but systems for feeding data into decision making.
- Build on what you're already doing.
- Use students' classroom work, evaluated by faculty, as a valuable source of information about learning.
- Keep it simple!

What Is Assessment?

Assessment is the systematic collection of information about student learning, using the time, knowledge, expertise, and resources available, in order to inform decisions that affect student learning.

Assessment as a Natural, Scholarly Act

Assessment is a natural, inescapable, human, and scholarly act. When we spend time teaching students how to shape an argument or solve an equation, we naturally ask, "Well, did they learn it?" Our academic training urges us to look for evidence to support claims, so when the college catalogue claims that students learn to be critical thinkers, we ask, "Well, do they?"

We're Already Doing Assessment

Assessment is so natural we have been doing it all along. Whenever a department or program says, "Students' senior capstone projects showed that, as a group, they are not doing well on X. Maybe we could . . ."—that's assessment. It happens all the time in responsible departments and programs.

Assessment as a Reform Movement

Assessment is a powerful national reform movement. The movement draws from public dissatisfaction with the perceived shortcomings of college graduates. Proponents of assessment believe that higher education should examine what students have learned, not just what the institution or department did that supposedly resulted in learning. The movement has become a mandate, imposed by accreditors and by some boards and state legislatures. Issues of accountability and public disclosure have become conflated with assessment (Ewell, 2004). It's a complicated scene. To follow the national movement, consult Ewell (2008) and the pages of the monthly newsletter *Assessment Update*, especially the columns by Ewell (www.interscience.wiley.com).

Movements and mandates may present both opportunities and dangers. Faculty often voice fears that appropriate faculty control over what is taught and how it is tested will be curtailed; results of assessment will be used irresponsibly; standardized tests will drive instruction; the goals of higher education will be dumbed down to what is measurable only in a narrow sense; higher education will be held responsible for things it can't control, such as the students' previous education or their lack of motivation; or educators will be forced to create costly and time-consuming bureaucratic systems that comply with accreditors' demands for assessment but that do not really result in improved student learning. These are real dangers. But the answer is not to ignore assessment, resist it, or leave it to others. Instead, *we must improve our assessment systems so that they help us enhance student learning, draw upon the best aspects of academic culture, and are sustainable in terms of time and resources. Then we need to explain our assessment systems clearly and without arrogance to our various constituencies.* I believe that we and our students can profit from assessment while minimizing the dangers. The purpose of this book is to show how.

The Three Steps of Assessment

The good news is that accreditors ask us to follow three steps that are natural and scholarly:

1. *Goals.* What do we want students to be able to do when they complete our courses of study? (Goals may also be called "outcomes" or "objectives." Issues of language are discussed later in this chapter.)

2. *Information.* How well are students achieving these goals, and what factors influence their learning? (Information may be called "measures" or "evidence.")
3. *Action.* How can we use the information to improve student learning? (Using the information may be called "closing the loop.")

Sometimes an additional step is added between 2 and 3: identifying where in the curriculum the goals are addressed (sometimes called "curriculum mapping"; see example in Appendix A). This step is not assessment per se, because it focuses on what the institution or department does to bring about student learning, not on what the students learned. Nevertheless, curriculum mapping is useful to identify goals that are not being consistently addressed. The three steps of assessment are discussed in detail within this chapter and the other chapters in this book.

Classroom Assessment and Program Assessment

Classroom assessment takes place within the confines of a single class. The instructor examines student work, talks with students about what worked for them, and then makes changes to his or her pedagogy or classroom activities.

Program assessment involves the department, program, general education, or institution examining student learning within those larger arenas and then taking action. For example, a department may examine a sample of capstone research projects from its senior undergraduate majors, as well as results from a senior student survey, in order to determine where the department can improve students' learning within the program as a whole. A general education committee may examine student work from a sample of general education courses, not to evaluate each teacher's performance but to assess how well the general education program as a whole is meeting its goals.

The End of Assessment Is Action

The goal of assessment is information-based decision making. To put it another way, *the end of assessment is action.* Assessment helps the organization determine how well it is achieving its goals and suggests effective steps for improvement.

That means you should conduct assessment for yourselves and your students, not just for compliance with accreditors. You don't need to build a whole superstructure of assessment bureaucracy; it's much more important to incorporate good assessment into all the institution's core decision-making processes that are already in place: departmental

decision making, committee deliberations, administrative policies, budgeting, and planning. You don't need to collect data you don't use; it's much more important to collect a small amount of useful data than to proliferate data that sit in a drawer or on a computer file. If you are collecting information you are not using, either start using it or stop collecting it. Instead of focusing on compliance, focus on the information you need for wise action. Remember that when you do assessment, whether in the department, the general education program, or at the institutional level, you are not trying to achieve the perfect research design; you are trying to gather enough data to provide a reasonable basis for action. You are looking for something to work on.

The Most Common Actions Resulting from Assessment

Three common actions that result from assessment in the department, in general education, and in the institution are these:

1. Changes to curriculum, requirements, programmatic structures, or other aspects of the students' course of study
2. Changes to the policies, funding, and planning that support learning
3. Faculty development

Sometimes the first action from an assessment is to gather additional information.

Pitfalls of Assessment

Common pitfalls of assessment include

- Mere compliance with external demands
- Gathering data no one will use
- Making the process too complicated

Section Summary

- Assessment is a natural, scholarly act, asking, "Are students learning what we want them to?" and "How can we better help them learn?"
- Assessment is also a national movement that poses both potential dangers and great promise for improving student learning.
- Assessment has three steps: goals, information, and action.
- The purpose of assessment is informed decision making.
- Assessment can go wrong when it focuses on compliance or on complex data gathering without using the information for decision making.

Concerns About Assessment

Aren't Grades Assessment?

Yes. But grades by themselves have limited use for program assessment. A department might know that the average grade on student senior research projects was 3.6, but that doesn't tell them much. It's not enough to say that we know students learned X if they got a grade of C or better in such-and-such a course. Instead, the department needs more specific, diagnostic information: students were strong in X and Y, but weak in Q and R. That detailed information tells the department what to work on. Such detailed information may emerge as faculty are grading student work, but then it must be aggregated and analyzed at the department or general education level, as each chapter in this book explains.

Sometimes grades can be used as a red flag. Especially, departments may want to monitor the grade distribution in introductory courses.

> Example: *Uncomfortable with the proportion of D and F grades and withdrawals from the introductory General Chemistry course at the University of Notre Dame, faculty members, led by Professor Dennis Jacobs, began a more extensive assessment. Faculty analyzed students' performance on the common chemistry exams and students' math Scholastic Aptitude Test (SAT) scores; they conducted interviews and focus groups with students; and they examined the research literature on how students most effectively learn in science. The grades were a red flag; the faculty used other data to expand their understanding of what was happening. Their findings and actions led to significant improvement in student learning (Jacobs, 2000, and Jacobs's Web site at www.nd.edu/~djacobs).*

How Can We Assess Complex Learning?

Assessment can and should be applied to the learning that the department, program, or institution most values, including the inclination to question assumptions, sensitivity to poverty and injustice, scientific literacy, the ability to work effectively with people of diverse backgrounds and cultures, or the development of ethical reasoning and action (for one list of liberal learning outcomes, see www.aacu.org/leap/vision.cfm).

We can't fully assess such ineffable qualities, but we can get indications. We are not caught between "objectivity" (in the sense that all judges of a student performance will agree on its quality) and "subjectivity" in the sense of individual whim. Between those two poles stands informed

judgment of work in our fields. As professionals, we assess our colleagues' work all the time. Assessing students' work is part of that responsibility. In assessing student work, not all judges of a single piece of student work will agree on its quality, but that's how disciplines move forward. If raters disagree, assessors can use established methods: take the average score, ask another rater to break the tie, or have raters discuss the student work to see whether they can come to agreement.

To get indications about how well our students are achieving ineffable goals, we must rely on student work or student actions that may offer only a small window into the ineffable quality. For example, suppose you want students to develop "ethical reasoning and action," which is one of the essential liberal learning outcomes identified by the LEAP (Liberal Education and America's Promise) project of the Association of American Colleges and Universities (www.aacu.org/leap/vision.cfm). To assess whether your students are developing this quality, you might rely on two methods:

1. Ask them in surveys whether they believe your program helped them develop ethical reasoning and action.
2. Evaluate something they do.

Under these two headings, many options are available. For example, Gelmon, Holland, Driscoll, Spring, and Kerrigan (2001) compare and contrast a variety of methods for assessment of aspects such as "awareness of community" and "sensitivity to diversity" that may result from students' service learning.

> Example: *Columbus State Community College faculty asked students to write about a scenario; the writings were evaluated for ability to "value diversity" (Hunt, 2000).*

• • •

> Example: *The United States Military Academy assesses students' "moral awareness" through analysis of classroom work, student surveys, and employer feedback (Forest and Keith, 2004).*

If your accreditor requires that you construct "measureable objectives," you don't have to abandon your high goals; you just have to identify, within the larger goals, some more concrete goals or "objectives" that can be applied to the student work you will analyze. For example, some departments, addressing the overarching goal of students' ethical behavior, choose to measure whether students follow the ethical principles of the discipline as they conduct research and write papers. Others choose to

measure whether students can identify and discuss ethical issues in case studies or scenarios. These measures do not address the entire concept of "ethical behavior," but they give indications about whether students are achieving aspects of your broader goal.

Can Assessment Be Applied to Online and Accelerated Learning?

Yes. Assessment gathers information about student learning, no matter what the pedagogy or mode of communication. If students are weak in a certain concept, the remedy may be somewhat different in an online course than in a face-to-face course, but the basic assessment process is the same.

Does Assessment Violate Academic Freedom?

The Association of American Colleges and Universities' Board of Directors Statement on Academic Freedom and Educational Responsibility (2006) directly addresses the issue of assessment and academic freedom:

> There is, however, an additional dimension of academic freedom that was not well developed in the original principles, and that has to do with the responsibilities of faculty members for educational programs. Faculty are responsible for establishing goals for student learning, for designing and implementing programs of general education and specialized study that intentionally cultivate the intended learning, and for assessing students' achievement. In these matters, faculty must work collaboratively with their colleagues in their departments, schools, and institutions as well as with relevant administrators. Academic freedom is necessary not just so faculty members can conduct their individual research and teach their own courses, but so they can enable students—through whole college programs of study—to acquire the learning they need to contribute to society.

Does Assessment Violate Student Privacy?

You do not need permission from your institutional review board (IRB) for normal assessment procedures described in this book. Federal policy exempts "(1) Research conducted in established or commonly accepted educational settings, involving normal educational practices, such as (i) research on regular and special education instructional strategies, or (ii) research on the effectiveness of or the comparison among instructional techniques, curricula, or classroom management methods. (2) Research involving the use of educational tests (cognitive, diagnostic, aptitude,

achievement), survey procedures, interview procedures or observation of public behavior." *However*, you must be sure that individual students cannot be identified and that they would not be harmed by disclosure of their responses outside the research (U.S. Department of Health and Human Services, 2008).

If you are conducting assessment for a grant-funded project or for publication, or if you have any doubts, check with your IRB. You are free to use the student permission form in Appendix B of this book, which I developed for a published study that used students' classroom work, journals, surveys, interviews, and classroom observations to explore teaching and learning in undergraduate introductory religion courses (Walvoord, 2008). This form passed the human subjects review boards of sixty-two institutions, ranging from public research-intensive universities to small private liberal arts colleges. Another example—a consent statement used for a grant-funded program to study critical thinking at Washington State University—appears in Maki (2004, p. 197). For a discussion of special issues in medical schools, see Brainard (2004).

Will Assessment Be Used in Tenure and Promotion Decisions?

Assessment is an evaluation of student learning to determine what faculty as a whole can do to improve that learning. A wise institution keeps the focus on collective action, not on individual blame. Keep a barrier between personnel decisions, which require administrative action and which protect personal privacy, and, on the other hand, program assessment, which requires collegial action by the department or institution. If assessment reveals a problem that can only be addressed by getting rid of a faculty member or changing his or her individual teaching practices, move that problem to the personnel side, and choose another problem for departmental assessment action. My usual advice is not to use student course evaluations both for personnel decisions and for program assessment. Develop a different instrument for programs—one that asks students how the program as a whole has contributed to their learning, or how it could be improved. Later chapters contain more details about these methods.

That said, an individual faculty member may use information about student learning as part of evidence for tenure, reappointment, or promotion. Evidence of learning can balance low student evaluations, for example. But the opposite is also true. Evidence of inadequate student learning in one's class ought to galvanize the teacher and the department for appropriate action. That action must be collegial and supportive, just as it optimally is when a faculty member is not producing sufficient

research. The truth is that assessment brings to teaching a level of accountability that was not always present before. The issues are complicated. The goal, however, is that assessment should be a collegial effort aimed at working together, as a team, to improve student learning.

Student Learning Is Affected by Factors Beyond Faculty Control

True, it is. But faculty, departmental, and institutional decisions do affect learning. A wise assessment program focuses on those factors you can control. For some audiences and purposes, you may also want to gather information about factors beyond your control, such as students' incoming skill levels or the number of hours they spend at their jobs, in order to establish students' beginning points or to present a fair picture of the context for student learning in your institution.

Section Summary

- Grades are only minimally useful for assessment; much more important are evaluations of the strengths and weaknesses of student work.
- Assessment can address complex learning.
- Concerns about assessment must be handled thoughtfully, but they need not be roadblocks to effective assessment.

Benefits of Assessment

Faculty sometimes ask, "Is there any research that shows that student learning improved as a result of assessment?" Yes and no. It would be impossible to design a research study that investigated whether, overall in the United States, student learning has improved because instructors and institutions are now being asked to do assessment in more explicit ways than before. There are too many variables; the meaning of "student learning" in that question is too broad; and assessment has always been occurring, even if not called by that name. Further, it's not the assessment itself that leads to improvement, but the action taken. The child's growth comes not from weighing her, but from feeding her.

What *can* be said, however, is that in countless individual instances, in departments and institutions of all types, assessment has been well used as a tool to help faculty and institutions make decisions that affect student learning, and that in some of those cases, the department, program, or institution has collected evidence that the new actions appear

to have enhanced students' learning. In 1993, Banta surveyed assessment coordinators at 115 institutions to collect their stories of how assessment findings had been used for improvement. She bemoans the lack of controlled longitudinal studies to track improvement in learning (one could still bemoan that lack), but some of the individual cases, even at that early time, offer documentation of improvement after assessment-informed actions. More recent collections of case studies include Banta, Jones, and Black (2009); Bresciani (2007); the journal *Assessment Update* and the edited collections of *Assessment Update* articles published by Jossey-Bass; and case histories of institutions that have improved learning and cost-efficiency in basic courses under the National Center for Academic Transformation (www.theNCAT.org).

Finally, "Does assessment improve learning?" is the wrong question anyway. We have to make decisions about curriculum, policies, resources, and pedagogies. We can make those decisions *with* information about student learning or *without* it. People have always sought information to inform their actions. In higher education, assessment is the answer to the latest educational fad, because it asks that we gather information about how well students are learning and that we use that information to inform our actions rather than just go along with what's currently in vogue. Assessment gives us another basis for action besides what people think would work, what other people do, or what's in someone's self-interest or convenience.

The right question is, "Since of course we want viable information to inform our actions, what information do we need, and how can we effectively gather, interpret, and use it?" In times of severely limited resources, we need assessment more than ever. When money is tight, time is stretched to the limit, and we're at our wits' end, we should not be saying, "We don't have time or resources to do assessment." Instead, we should be saying, "Let's use well-conducted assessment to help us achieve our aims most efficiently in this difficult time."

Section Summary

- We naturally want good information about student learning to inform our actions.
- Assessment, when well conducted, can help us gather that information and use it effectively.
- When used effectively, assessment can lead to improvement in learning.

Communicating About Assessment

We not only must do assessment well, but we must communicate about our assessment to accreditors and other external audiences. Each chapter of this book discusses how to write assessment reports and self-studies; this section lays down the general principles that guide such communication.

Assessment Is Cross-Cultural Communication

Assessment is an exercise in cross-cultural communication among various segments of the academy and between the academy and those it serves. Bergquist and Pawlak's *Engaging the Six Cultures of the Academy* (2008) describes the cultures that coexist on most campuses and that influence how "assessment" is named, perceived, and used. I would place the assessment movement primarily within the "managerial" culture, which values the articulation of goals and objectives for students' learning, purposeful planning to achieve those goals, and the use of data to evaluate the achievement. Assessment also contains aspects of the "developmental" culture, which seeks to further the intellectual and personal development of both students and faculty and which relies on institutional research for information about this development. Many faculty, on the other hand, are part of the "collegial" culture, marked by high value on disciplinary research, high insistence on faculty autonomy, and ambiguity toward accountability for student learning. Each culture, however, has aspects that overlap the others, and individuals may function as members of multiple cultures. People in all cultures may care deeply about student learning, and each culture has ways of conducting assessment, though they may not use the word. Good communication about assessment can be built by looking for common ground, addressing various audiences with language that is accurate and familiar to them, and being honest.

Who Needs to Know *What*, for *What*?

It is easy to focus solely on accreditors as the audience for assessment, but assessment information may also be useful to potential students, donors, the general public, legislature, board, and others. The most important audiences may be your own students, faculty, and administrators. Audience and purpose should direct your choice of assessment methods: Questions such as "Should we use portfolios?" or "Should we use a standardized test?" can be answered by determining "Who needs to know what, for what?" Appendix C is intended to help you plan for your audiences.

Accreditors Are Not the Enemy

The regional accreditors are subject to federal oversight; from many directions, they feel pressure to be tough on assessment. Accreditors' staffs and funding are stretched thin. Critics have proposed that they be abolished and replaced with a single national accrediting agency staffed by educational measurement experts, instead of the current visiting teams comprised of faculty and administrators from other campuses. If we want to keep our peer review system, it's in our best interest to help our regional accreditors do their jobs. Treat them not as the enemy, but as colleagues and collaborators.

Use Self-Reflective Analysis

An accreditation study is not a public relations piece but a candid analysis. When I coordinated the writing of the self-study for my own institution, every section had a description of our actions in that particular area, followed by a section headed "strengths," a section headed "weaknesses," and a section headed "future plans." If you try to cover up weaknesses, or puff up your report, the accreditors are tempted to turn into "gotcha" police. Instead, you want to enlist these visitors as colleagues in candid conversation about the strengths and weaknesses of your current system and what can best be done to improve it.

Section Summary

- In communicating about assessment, consider the needs of your audience, recognizing that you may have to cross cultural lines.
- Ask "Who needs to know what, for what?"
- Adopt a self-reflective posture. Seek collegial conversation with those who are asking you for assessment.

General Guidelines for the Three Steps

This section presents some general guidelines that underlie the more specific advice that appears in later chapters for institutional leaders (Chapter Two), department chairs (Chapter Three), and general education (Chapter Four). I have arranged the guidelines under the three steps—goals, information, and action—but the purpose here is not to give a complete guide to the three steps themselves but to present those general principles that are common to all levels of assessment and that I do not want to repeat in each following chapter.

Guidelines for Establishing Goals

The first step of assessment is to establish learning goals. This section lays down some common principles; each successive chapter discusses how to establish goals in its particular domain.

Format: "Students Will Be Able to . . ."

Goals must be in the "students will be able to . . ." format. Here are some goal statements that are *not acceptable* for this purpose (though they may be perfectly fine statements for other purposes):

- The curriculum emphasizes X, Y, Z.
- The institution values X, Y, Z.
- The institution prepares its students for X, Y, Z.
- Students are exposed to X, Y, Z.
- Students participate in X, Y, Z.

Terms: Goals, Objectives, or Outcomes?

I use the term *goals* throughout, but in various settings, you will find other terms such as *objectives* or *outcomes* (student learning outcomes are sometimes referred to as "SLOs"). These terms are used inconsistently in the literature, so don't get hung up on the distinctions. Trying to get a whole faculty to understand and consistently employ a particular distinction among these terms may be futile. I suggest choosing one term, such as *goals*, with the understanding that the goals must be stated at various levels of generality. If your accreditor, board, or system is using any of these terms with a specific meaning, good communication practice would suggest that you use their terms when you write for them.

Levels of Generality

You will state the goals at various levels of generality. For example:

- *Institutional level.* Students will communicate effectively in writing to a variety of audiences.
- *Department/school/college level.* Students who complete the business major will communicate effectively to professional and lay audiences, using the common business formats.
- *Course level.* When they complete this finance course, students will be able to write such-and-such kinds of financial reports.

Subsequent chapters discuss goals for the institution as a whole (Chapter Two), for departments (Chapter Three), and for general education (Chapter Four).

Section Summary

- State goals as "students will be able to . . ."
- The terms *goals, objectives,* and *outcomes* are used inconsistently in the field; keep it simple; meet the needs of your audiences.
- Goals must be stated at different levels of generality for different levels of assessment.

Guidelines for Gathering Information

The second step of assessment is to gather information about how well students are achieving the goals. Again, subsequent chapters show specifically how to accomplish this step in the institution, departments, and general education. Next are principles for two common types of direct measures: standardized tests and classroom work. Each has its own benefits and drawbacks.

Standardized Tests

Assessment per se does not require standardized tests; it asks for a sensible combination of measures that will yield useful, actionable information about student learning, including some direct measures. Standardized tests offer scores that can be compared to national samples, and they test students against national standards of performance. Seeing how your students perform against a national standard may provide bragging rights or it may be a shock. Both can be useful. However, if your students score low on the test, you have no way to improve their scores unless your faculty are willing to teach what the test is measuring. In addition, standardized tests may present other methodological problems. If the test does not count toward regular academic work, students may not do their best. Getting a meaningful sample of students to take the test can be difficult. Tests are expensive. Chapter Two discusses in more detail the institutional decision about administering a national standardized test and about joining collaborative agreements, such as the Voluntary System of Accountability, to make scores public.

Classroom Work: Samples and Portfolios

Classroom work, like standardized tests, has advantages and limitations. One advantage is that classroom work is already being examined by faculty as part of the grading process, so evaluating it for assessment purposes can be time efficient and relatively inexpensive. Classroom work carries a grade, so students may be more highly motivated than if they

took a standardized test that did not count in their grades. Classroom work reflects what students actually are taught at your institution, rather than what standardized test constructors think should be taught. Faculty may be more invested in their own analysis of students' work than they are in the results of some standardized test. And classroom work can be evaluated in many ways over time, to yield many kinds of insights.

On the other hand, classroom work does not yield scores that can be compared across institutions. And it takes careful work to aggregate classroom work or portfolios so that the results can be used for action in the department or program, general education, or the institution as a whole. The next sections of this chapter discuss aspects of using classroom work that are common to all situations. Succeeding chapters discuss in more detail how classroom work may be used for institutional, departmental, and general-education assessment.

The final chapters of my and Anderson's *Effective Grading* (2010) also discuss how to aggregate and use classroom work for assessment in grant-funded projects, departments, and general education. Also useful is Banta's (2003) collection of *Assessment Update* articles about portfolio uses, cases, scoring, and impact, and Zubizarreta's (2009) useful book on portfolios, including their uses in classrooms.

Gathering Student Classroom Work. Depending on your assessment questions, audiences, and purposes, you should give careful thought to the amount and type of student work you need to collect. Here are some common selections:

- Penultimate work from a course toward the end of the students' course of study. For example, a capstone research paper, concert, theater performance, or internship. Answers questions such as, "What are students' strengths and weaknesses at the end of our program?" An excellent starting point for any assessment program. From this work, select a weakness to work on, and, if necessary, gather further information from earlier student work.
- Pre-post: Sample of student work at the beginning of their course of study and at the end. Answers questions such as, "What is the 'value added' for students in our program of study?"
- Portfolios of student work (a *portfolio* is multiple pieces of one student's work completed across time). Answers questions such as, "How do our students develop?" or "How well can students exhibit particular skills in a variety of settings?"

You can gather student work (either samples or portfolios) in two ways:

1. Ask a sample of faculty to submit copies of their students' work and/or to analyze their students' work.

Example: *The Office of Institutional Research identifies each year a random sample of general education courses that are focusing on, say, critical thinking. The chosen course instructors submit their student papers for an assignment that asked for critical thinking. These papers are evaluated by a group of faculty readers who produce a report that is disseminated among general education faculty and used at the institutional level for action to improve the general education program. A process similar to this one was used at Johnson County Community College in Kansas (Seybert and O'Hara, 1997).*

• • •

Example: *In a department, instructors of 400-level courses that ask for research papers are asked to submit rubric scores or analyses of the strengths and weaknesses of their own students' work, measured against one or more departmental learning goals. These reports are aggregated to inform departmental decision making.*

2. Ask a sample of students to submit their work.

Example: *A random sample of students are contacted and asked (or paid) to submit copies of their work. Interviews or other information may also be gathered to provide a deeper set of data.*

• • •

Example: *Software may be used to collect student work. As students submit their work online, they give permission for their work to be used as part of a sample for assessment. Assessment projects can then select a sample from the online work.*

• • •

Example: *Students in a capstone course compile portfolios as part of their coursework. The portfolios are used for department or general education assessment.*

Establishing Criteria for Evaluating Student Work. When you evaluate student work, you need a set of criteria or a set of questions. You can use a rubric or an alternative method.

The rubric is a format for expressing criteria and standards. The advantage of a rubric is that it disaggregates various qualities of the students' work. Thus it is diagnostic; it helps you see what to work on. Instead of "Students' average grade on the capstone project was B+," a rubric helps you say, "On the capstone project, students' strengths were P and Q, and their weaknesses were X and Y."

Rubrics can be minimal or full. A minimal rubric simply lists the traits on which the evaluation will be based, and it indicates a scale, but without describing the student's performance at the various levels of the scale (Exhibit 1.1).

Minimal rubrics may help a group of faculty identify what they value, but you won't get very high percentages of interrater reliability (instances where readers give the same paper the same score). You also won't get much detail about what is going wrong if, for example, students score low on "organization."

A full rubric addresses these problems. It describes each level of performance. Exhibit 1.2 is one section of a rubric developed for assessing essays in which students were to take a stand on a debatable issue about a work of literature (full rubric is included as Example 1 in Appendix D).

A rubric may be used as the basis for a grade by weighting the various traits (for example, thesis counts 25 percent, support counts 40 percent, and so on), or the rubric may simply be shared with students as a guide to their work, not necessarily correlating numerically with a grade. *Effective Grading* (Walvoord and Anderson, 2010) discusses in detail how to share

EXHIBIT 1.1
Minimal Rubrics

Option 1. Naming the Traits					
Thesis	5	4	3	2	1
Organization	5	4	3	2	1
Etc.					
Option 2. Describing the Top Performance					
Thesis is clear, debatable, complex, and creative.	5	4	3	2	1

EXHIBIT 1.2

Selection from a Full Rubric for Essay of Literary Analysis

5	4	3	2	1
Thesis: The thesis of the paper is clear, complex, and challenging. It does not merely state the obvious or exactly repeat others' viewpoints, but creatively and thoughtfully opens up our thinking about the work.	The thesis is both clear and reasonably complex.	The thesis is clear, though it may be unimaginative, largely a recapitulation of readings and class discussion, and/or fairly obvious.	Thesis is discernible, but the reader has to work to understand it, or the thesis seems to change as the essay proceeds.	Thesis is irrelevant to the assignment and/or not discernible.

grading criteria with students and how to relate rubrics to grades. A basic principle is that, if you're using a rubric, it should be shared with students before they begin the assignment or test.

Rubrics may be constructed either by individuals or by groups such as a department or a general education committee. Steps for constructing rubrics and answers to frequently asked questions about rubrics can be found in my and Anderson's *Effective Grading* (2010), which places rubrics in the larger context of course planning and pedagogy. Within a more narrow context, Stevens and Levi (2005) offer a guide to constructing and using rubrics. For examples of how rubrics have been used for program assessment, see Case Study 2 in Chapter Three of this volume, the case studies in Chapter Twelve of *Effective Grading*, and many of the case studies in Banta, Jones, and Black (2009), in Bresciani (2007), and in the various collections of *Assessment Update* articles edited by Banta and published by Jossey-Bass (Banta 2004, 2007a, 2007b; Banta and Associates, 2003).

While I was teaching hundreds of faculty to construct rubrics, I have found that some faculty take to rubric construction immediately, and ask, "Why didn't I start using rubrics years ago?" Other faculty, who may be equally smart and equally good at teaching, find it hard to bend their thinking into a rubric format. If you're not comfortable with rubrics, then use a list of criteria and analyze the students' work for strengths and weaknesses. As you read each paper, make a list of the strengths and weaknesses it exhibits in meeting

the criteria. (If you're just beginning to shape criteria, let the criteria statements grow and change as you do this.) At the end, aggregate the lists of strengths and weaknesses to find those most common to the papers as a whole.

Example: *A group of philosophy faculty read a selection of senior student research papers, jotting down notes about the papers' strengths and weaknesses as they read, and then identifying the overall strengths and weaknesses of the group of research papers as a whole. Faculty took their written notes to a meeting of all the readers, where, in conversation, they identified the strengths they wanted to celebrate and one student weakness they wanted to work on. No rubrics were used, but the process yielded action based on careful faculty analysis of student work.*

• • •

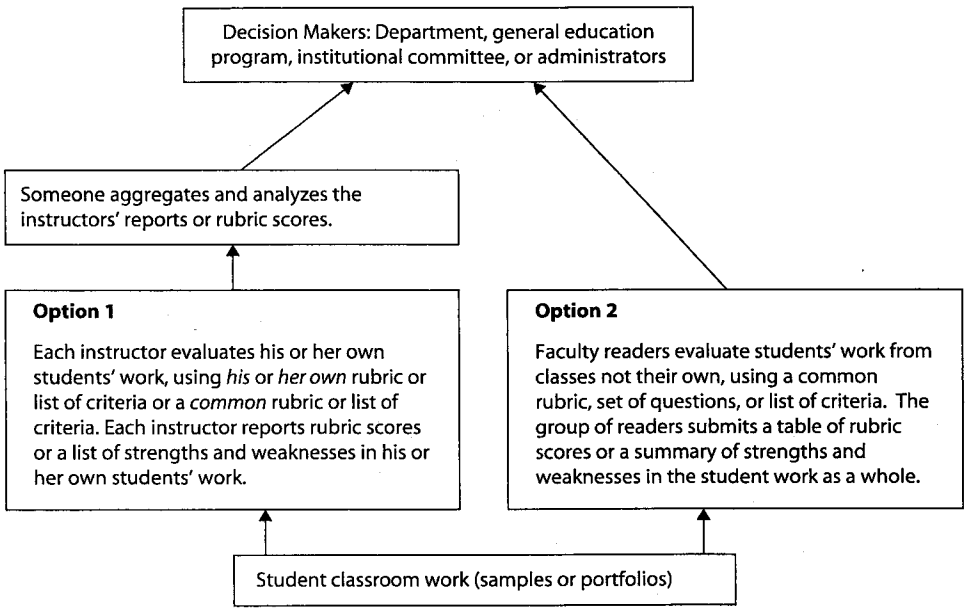
Example: *At one institution, a sample of fifty students from various disciplines were paid to assemble portfolios containing selected pieces of their work across four years (their first paper from their first semester college seminar, the last paper in that seminar, what they judged to be the best paper of their second semester, the first paper they wrote in a course in their major, and their best paper from a senior course in their major). These were combined in the portfolios together with annual interviews conducted by a researcher, annual surveys completed by the student, and transcripts of the student's coursework and grades. A group of ten faculty were offered stipends during the summer, each to read ten complete portfolios. (Thus, each portfolio was read by two faculty members.) The faculty readers were asked to address a set of questions that the group had determined ahead of time, but they were also asked simply to note what struck them or surprised them about the portfolios. What themes and patterns emerged? What were points of contrast and comparison? Then the group of ten faculty met for broad-ranging discussion about what they had found—a discussion that was eventually narrowed to identify several areas for further investigation by more focused techniques. This strategy employs the broad "What's going on here?" question used by qualitative researchers as they initially survey a broad range of data.*

Aggregating and Analyzing Student Work. Classroom work (samples or portfolios) can be aggregated for use by the department, general education, or the institution in either of two ways (Figure 1.1).

In Option 1, the instructor (piggy-backing on the grading process) prepares a report of students' strengths and weaknesses or rubric scores on

FIGURE 1.1

Evaluating Student Classroom Work: Two Options



one assignment or a portfolio of assignments for the class as a whole or for some predetermined sample of class members. These individual instructor reports are then aggregated. For example, one might examine fifty faculty members' reports to discover which student weaknesses are most often mentioned. This is the least time-consuming method of gathering classroom data because the papers are read only once—by the instructor. Everything else is based on the instructor's report or rubric scores.

In Option 2, the instructor may play no role or may serve only as the collector of student work, not the analyzer. The pieces of student work or portfolios are read by a separate group of readers (usually faculty but sometimes graduate students), who prepare a single report for the department or general education. This method introduces external eyes, but the faculty readers may not understand what they are reading. Moreover, Option 2 is more time consuming, because each piece of student work is read twice—once by the instructor for a grade, and again by the faculty readers for assessment.

If you are evaluating pieces of work or portfolios that are *similar* to one another in type and discipline, you may use either of the two options. However, if the student work is *different* (for example, you are trying to

evaluate critical thinking or writing skills from student work in disparate courses such as history and physics, or work of disparate types such as the case study and the research report), it may be very difficult and time consuming to follow Option 2, the common rubric used for all the student work. Before you embark on constructing a common rubric for varied types of student work, ask yourself why you need a single set of rubric scores. Read the case study of Tompkins Cortland Community College (Cameron, 2009) for a description of the five years of arduous work involved in constructing and using a single rubric for disparate work, and the ongoing difficulties with interrater reliability, as well as the benefits of this process for faculty.

You don't have to use a single common rubric to evaluate disparate types of student work. An alternative is to construct (or adapt from elsewhere) a general rubric or a set of criteria or learning goals to be used only as a *guideline* and let disciplinary groups construct their own discipline-specific or assignment-specific rubrics and/or lists of criteria based on the general guidelines. Then you take the individual rubric scores or lists of strengths and weaknesses, and aggregate them, looking for common themes or for difficulties that are most often mentioned. Such an aggregation can serve as the basis for action. If many of the individual rubric scores or lists of strengths and weaknesses show difficulty with using appropriate online sources, work on that. If they show difficulty in considering alternative solutions/explanations/arguments, work on that. For general rubrics that you might be able to adopt or adapt, follow the work of the Association of American Colleges and Universities (www.aacu.org) and the Teagle Foundation (www.teagle.org).

Section Summary

- Standardized tests are not required for assessment. Use them if they yield information you can act on.
- To state criteria and standards for student work, you can use rubrics or alternatives such as lists of criteria.
- To evaluate student work, you can have individual instructors evaluate their own students' work and then aggregate those evaluations, or you can have a group of readers evaluate student work that is not their own.
- Using a single rubric for disparate work is difficult; an alternative is to use a general rubric or set of criteria as a guide, and let individual departments or instructors create their own assignment-specific versions.

Guidelines for Action

The third step of assessment is to act on the information—sometimes called "closing the loop." Each succeeding chapter discusses how to close the loop in various situations; this section presents general principles.

Identify Factors Affecting Learning

Knowing the strengths and weaknesses of student work does not necessarily tell you what to *do* to improve student learning. For that, you need information and hypotheses about what is affecting student learning. You will probably use two ways to collect this information and formulate hypotheses:

1. Rely on the research that is already available.
2. Do some research yourself.

The research that is already available tells you what factors are likely to be affecting your own students' learning, and what actions might help. For analysis of the institutional characteristics linked to student success, consult Kuh (2008), Kuh and others (2005a, 2005b, 2007), and Pascarella and Terenzini (2005). A study of church-related colleges is Braskamp, Trautvetter, and Ward (2006). For the teaching methods that research has found most successful for learning, consult Chickering and Gamson (1987), which is widely available on many Web sites. For a given situation, more specific research may be helpful. Then let the published research guide and complement your own investigation.

> Example: *Teaching the general chemistry course we mentioned earlier, which was trying to help a greater proportion of its students be successful in the course without lowering the standards, faculty consulted the research literature to identify possible causes of students' failure and to find teaching methods that had worked in other settings. The faculty also analyzed students' tests, interviewed students, and examined their SAT scores. The literature, as well as the department's own research, suggested that some students had not learned the problem-solving strategies that were necessary for college-level chemistry, and that the large lecture format was not very effective for these students. The research suggested that integrating small-group problem-solving and frequent graded homework problems would enhance students' problem-solving skills. The department implemented those strategies, and a greater percentage of students did better on the same final exam. (For the case history on which my short incident here is based, see Jacobs, 2000, and www.nd.edu/~djacobs).*

Are the Actions Working?

If you were to "close the loop" in a scientific way, you would not only use assessment information to inform your action, but you would then come back and examine whether your action was achieving the improvement of

student learning you had hoped. The chemistry department was able to do just that, by comparing students' scores on the common final exam, to see whether students in the experimental section did better than a matched group of students taught by the old pedagogies.

This type of reexamination to see whether a strategy worked can be effective in departments or in individual general education programs such as the required math or composition class. The scale is often small enough to track the impact of changes on student learning. However, suppose a general education program noted, from standardized test scores, student surveys, and/or faculty review of portfolios, that students' writing skills were not what the faculty had hoped. The institution consults the research literature, interviews students, examines how writing is being taught across its curriculum, and takes significant steps to address the writing issue. Could the institution, five years later, expect that these changes would result in an improvement in students' writing scores on a standardized test or on faculty-scored portfolios? The institution would certainly want to continue to track the test and portfolio scores, but there might be so many variables at work, so many other changes taking place over time, and such a large, complex population of students, that no significant difference in overall scores would appear. That doesn't necessarily mean the changes were worthless; it may simply mean that there are too many variables, and the nature of "writing" is simply too different in different disciplines and contexts. In that situation, one might investigate smaller settings: for example, take ten individual courses in various disciplines in which faculty adopted new pedagogies to enhance student writing. Investigate whether student writing improved in those courses, using samples of student writing in each course.

Further examples and details about "closing the loop" are presented in the chapters that follow.

Section Summary

- To inform your action, you need information and hypotheses about the factors that affect learning. Gather these from the published literature and your own investigation.
- If your context is too broad or complex to track whether specific changes are producing improvement in learning, you can take a sample of local situations, such as individual classes or programs, where improvement in learning can more easily be tracked.

Okay, So What Should We Do?

People often ask me for a specific list of "what we should do" for assessment. The lists below are not necessarily exhaustive, but they provide a framework for how each segment of the institution could do assessment in ways that are helpful to decision making, consonant with natural and scholarly processes, and acceptable to accreditors and other external audiences.

Classroom instructors could

- Articulate what we want students to be able to do when they complete our courses
- Gather information about student learning from our own classroom assignments/exams, and from surveys, focus groups, or conversations with our own students about their learning experience, and use that information to improve student learning in our own classes
- Be willing to bring that information to the department or general education program to be aggregated with other classroom information and used for decision making in areas where problems needs to be addressed at levels beyond the individual classroom
- Keep records of our assessment work and report that work as needed

Departments or programs could

- Articulate what students should be able to do when they complete each of our certificate or degree programs and our general education and service offerings
- Gather information from a sample of students' classroom work, from student feedback, and from other sources as relevant, and use that information for decisions and actions that affect student learning
- Be willing to bring our information to other decision-making committees, offices, or administrators, so it can be aggregated with other information and used for decision making.
- Keep records of our assessment work and report that work as needed

General education programs could

- Articulate what students should be able to do when they complete the general education curriculum
- Develop subgoals for individual general education programs such as the composition program, the math program, learning communities, or community-based learning
- Gather information from a sample of students' classroom work, from student feedback, and from other relevant sources, and use that

information to make decisions about general education curriculum and policies, and/or to offer faculty development
- Be willing to bring that information to other decision-making sites
- Keep records of our assessment work and report that work as needed

Faculty committees, governance bodies, and administrators could

- Support assessment efforts with resources, policies, and encouragement
- Ensure that the institution has a consistent, integrated assessment system that uses information about student learning for improvement at every level—department, general education, and institution as a whole
- Keep records of assessment work and report that work as needed

Chapter Summary

- The end of assessment is action.
- Pitfalls of assessment include merely complying with external demands, gathering data no one will use, and making the process too complicated.
- Grades are only minimally useful; instead, you need diagnostic information about students' strengths and weaknesses.
- *Goals, objectives, outcomes*—don't sweat the terminology. Just state "students will be able to . . ." at various levels of generality.
- Standardized tests may be necessary for accountability, but if you have a choice, examine your options very carefully. Consider using student classroom work.
- Students' classroom work is a valuable source of information about learning, if you evaluate it by rubrics or lists of criteria.
- It is very difficult to use a single common rubric for varied types of student work. Consider using the generic rubric as a guide, and letting individual programs or instructors generate their own assignment-specific rubrics.
- To act on information about learning, you need research and hypotheses about the factors that may affect learning.
- Do your best to track the results of changes you make. In large, complex contexts, you can choose a sample of classrooms or programs to show results of changes.
- What to do: Follow the three steps: goals, information, action. Put in place a sensible, sustainable assessment system that helps you improve student learning, then explain your system to accreditors.
- Keep it simple!