Barbara D. Wright

An Assessment Primer

Over the last five to seven years, assessment has evolved into a highly effective, user-friendly tool for strengthening undergraduate education. This article explains, in simple terms, how assessment may be undertaken, what purposes it serves, what methods are available, and what the emerging trends seem to be.

The word *assessment* is derived from the Latin verb *assidere*, which means "to sit beside" and help judge. As a great lover of words, I often think about their very literal origins and the permutations their meanings undergo; and this particular word, assessment, evokes pleasant images in my mind: a companionable scene in which an instructor is sitting with a student, going over a piece of work, and discussing its strengths and weaknesses; or a group of faculty, sitting around a seminar table, talking about the responses they are getting from a survey of their majors. Perhaps it's late afternoon in the fall, and a little grey and chilly outside, but the conversation inside is lively and warm.

I realize these are not the associations everyone makes when they hear the word "assessment"; it has meant very different things to different people at different times. For example, at the third national assessment conference, sponsored by the American Association for Higher Education (AAHE) in Chicago in 1988, some attendees went out to an improvisational comedy club. When the audience was asked to suggest subjects for a skit, someone in the group yelled out "assessment," and the troupe responded with a spoof on realtors and tax collectors. There is less confusion on our campuses, where, over the last seven years or so, assessment has evolved into a highly effective, user-friendly tool for helping all of us in higher education—faculty, professional staff and administrators, as well as students—to better understand what happens to the individual student, and how effective our programs or our institution as a whole may be.

This is not the place for a detailed history of postsecondary assessment. Others, like Peter Ewell in his exhaustive review of the literature, *To Capture the Ineffable*, or Ted Marchese and Pat Hutchings in

Watching Assessment, have already done a fine job of that. Suffice it to say that the movement has twin roots, and that it most likely would not be as strong if it were not for this double pressure to assess. Pressure is exerted both from within the higher education community and from without. Within higher education, there was that famous cluster of reports dating from the early 1980s that included, for example, *Involvement in Learning*, issued by the National Institute for Education in 1984, the Association of American Colleges' report *Integrity in the College Curriculum* (1985), and *To Reclaim a Legacy*, from the U.S. Department of Education in 1984. All of these set forth an agenda for reform of undergraduate education and more or less explicitly advocated assessment as a means to that end.

Meanwhile, there has been mounting external pressure to assess: from the federal government, from governors, legislators, and state departments of higher education, as well as from regional and professional accrediting agencies. And there has been a more diffuse but equally urgent call for better undergraduate education from society at large: from the employers of college graduates, from parents, from students themselves. The recent economic recession has focused a good deal of public attention on the need for the United States to train a more competent workforce, and to consider such questions as what that means and how it can be achieved. Americans are also concerned of late about the social needs of the United States, and about the contribution that a productive and educated citizenry could make to solving social problems. As educators, we cannot ignore this backdrop; it makes our work on campus with assessment much more than an interesting academic exercise.

What Is Assessment?

At AAHE, Ted Marchese and Pat Hutchings like to talk about assessment as "conversations"—conversations about what our students know and can do, and at what level; about how we can *know* what they are learning or how they are learning it; and what we can do to make the learning process go forward more effectively. I like that definition, but for many, particularly newcomers to assessment, it requires a bit of explication.

A more traditional definition of assessment might go something like this: Assessment is a systematic, ongoing process of (1) *setting goals* or *raising questions* about the education we provide; (2) *gathering information*; (3) *interpreting* it; and (4) *using* it to improve the effects of college on student learning and development. What can we say about this bare bones definition? First, it is important to note that assessment is an ongoing, iterative *process*, not a one-shot event. It consists of four basic steps, and once we have traveled through them, the sequence begins again. Second, the focus in this definition is on *improvement*—not methodology, or number crunching, or report writing, or accountability for their own sakes, but in the service of educational improvement.

Let us look a little more closely at the individual steps. The first is to define goals or raise questions about some aspect of the education we offer. We may say that a major goal of general education, for example, is for students to be able to write coherently and persuasively; or we may say

that a major goal of the education major is for students to pass the state teacher certification exam. But goals are often difficult to define explicitly, particularly when a campus is new to assessment. Faculty may then be more strongly moved instead by questions: How much writing *do* our students typically produce in the course of fulfilling their requirements? Is it enough? And of what quality? What *is* our students' current pass rate on the teacher certification exam?

The second step is gathering evidence that will demonstrate whether students are in fact reaching the institution's or the program's goals, evidence that will begin to answer the questions that have been raised. But gathering evidence requires assessors to think first about what *kinds* of evidence will be most useful and to the point, and then to figure out the most appropriate methods of gathering that evidence. The first thing that typically springs to mind at this point is giving a test; but test scores, particularly from off-the-shelf standardized, multiple-choice tests, may not provide the kind of campus-specific, detailed, and vivid information that will prove most useful for steps 3 and 4. If a campus is concerned about quality and quantity of student writing, for example, gathering examples of student work and surveying instructors about the writing assignments they give may be more to the point. If faculty in a program are concerned about student pass rates, they may want to talk directly to students who did well or poorly about their perceptions of *why* they performed as they did.

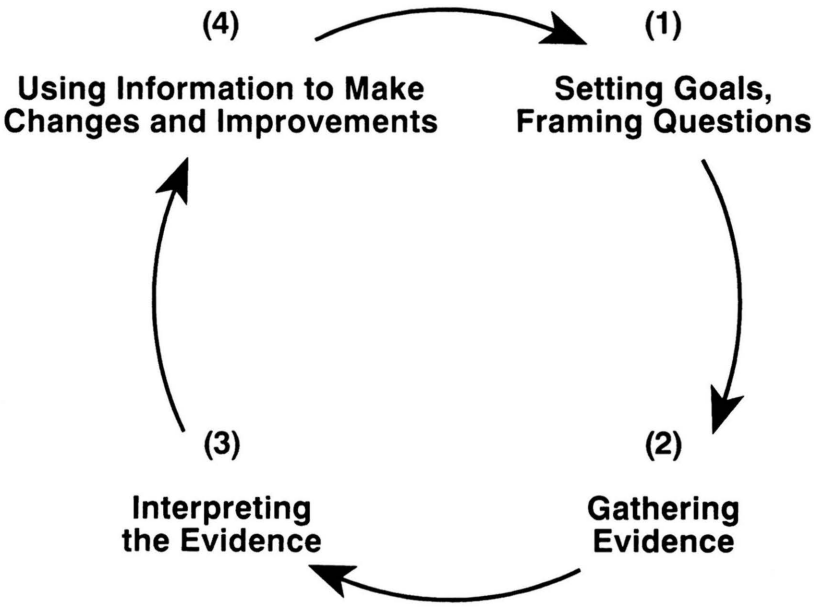
Traditionally, evaluators, accreditors, and others have talked about quantitative "measures" of effectiveness, but in my view it is essential to conceive of evidence—and of the methods used to gather it—more broadly. Measures are *one* form of evidence among many; but we also need to include other kinds of evidence, for example, samples of student research and writing, projects, videotapes of student performances, narrative descriptions of students' achievement and progress, and so on. There are distinct advantages to this more inclusive conception of assessment. For example, it becomes possible in this way to demonstrate education-related changes that are difficult or impossible to capture using standardized testing or surveys.

Detailed, provocative evidence is also far more useful than percentages or percentile scores alone in stimulating faculty discussion and substantive changes in teaching and learning. And that brings us to the third step: the evidence that has been gathered must be interpreted, data turned into information. The task of meaning making needs to include and respect the perspectives of various components of the campus community if it is to yield suggestions for change that are both credible and practical. Thus, for example, the quality and quantity of student writing must be viewed in relation to the level of skill students bring to campus, opportunities in courses or outside to learn to write well, the campus consensus (or lack of it) about the importance of writing, the amount of resources available to assist students, and the like. Students' pass rates must be interpreted in relation to entry-level knowledge and skills, the content and pedagogy of their course work, opportunities to teach and observe master teachers, students' perceptions of how their preparation helped or hindered them on the exam, and so on. The third step of interpreting the information,

more than any other in the four-fold process of assessment, illustrates how assessment is not a scientific experiment but a messy and challenging exercise in human judgment.

Finally, the evidence and the interpretation that has been made of it must be *used* for change and improvement. It is not enough simply to collect data, discuss it, and write a report or present recommendations. Assessment efforts need to result in concrete changes, which in turn can be studied as the entire cycle begins again. In the next iteration, the completeness and explicitness of goals, the appropriateness of information-gathering approaches, the accuracy of interpretations, and campus responses all can be revisited in light of greater knowledge of the educational process and greater experience in assessment. It is important, in other words, to think of assessment not as a linear process with a beginning and an end, but as a circle, a loop, that must be closed and then retraced again and again.

Figure 1



That, at least, is the basic idea. But one need not begin with step one. Setting goals is the logical first step, but not an easy one, and for philosophical, political, or other reasons it is sometimes impossible for an institution to reach consensus. If that is the case, it is better to respond pragmatically and begin with another step than not to begin at all. Faculty can, for example, start with the information-gathering stage. Working inductively instead of deductively, they can gather samples of student work, analyze it to figure out where their students are now, articulate

what seem to be the college's implicit, functional goals, and then decide whether these operative goals are appropriate and complete, or whether they need to be amended or augmented in some way.

My own university offers another example. At the University of Connecticut, something quite typical happened: a reform was instituted—that is, step 4 was taken—without much benefit of the other three steps. We developed a new general education program, but instead of the program being conceived in reference to a set of student outcomes goals or on the basis of detailed information about what students knew or could do, it was based on what faculty and administrators *thought* belonged in it, with an eye to potential turf battles and existing courses. The assessment committee's first task was then to produce an appropriate set of goals for the curriculum, after the fact, in relation to which students could be assessed. This wasn't exactly "classical" procedure, but it worked.

The four-step process meshes well with the definition of assessment as "conversation." Obviously, carrying out each of the four steps requires a tremendous amount of conversation involving the campus community to the fullest extent appropriate: faculty, students, campus professionals, administrators. Academics are notorious talkers, but assessment helps ensure *productive* conversations. It provides the framework so that such conversations happen routinely instead of by chance; it provides the grist for *informed* conversations; and it provides the procedures that turn good conversation into sensible action.

What Does Assessment Look At?

There are five basic domains for assessment of student learning outcomes: (1) basic or entry-level skills; (2) college-level skills; (3) general education; (4) the major and vocational or professional training; and (5) student development. We are most familiar with the basic skills assessments that have been used for years to identify students who require remediation in such areas as reading, writing, or math. These are skills that we expect students to acquire in high school and bring with them to college. College-level skills, in contrast, include such things as more sophisticated writing and math skills, and higher-order thinking skills like problem solving or critical thinking. Development of such skills is often a major focus of general education programs, along with more content-focused aims such as exposing students to the basic questions and methods of inquiry of the traditional academic disciplines.

But the definition of general education varies enormously from institution to institution: in the balance it strikes between skills development and content mastery; in its disciplinary or interdisciplinary nature; and in the breadth or narrowness of students' choice, ranging from tightly required core courses to flexible menus including dozens or even hundreds of courses. General education is one of the most important but most difficult places to begin assessment. The major, on the other hand, is a natural and relatively easy place to begin. It is an area in which faculty are invested and to which they are committed, one in which they feel expert, the one in which they have the clearest ideas about what students should

be able to do at the end of a course of study. It is also the academic area to which students are often most committed.

However, our expectations for students include more than mere knowledge and skills. We want them to develop certain habits of heart and mind, to embody values, to pursue interests, to exercise responsibility in their public and private lives. All that falls under the rubric of student development. It is as central to the mission of higher education as it is difficult to assess. Despite its close tie to academic learning, faculty often feel uncomfortable at the prospect of assessing student development, and it has often been viewed as the responsibility primarily of student affairs personnel.

At an institution undertaking a comprehensive effort, assessment usually includes these five domains, but it may also go beyond them. For example, an institution with very strong professional programs may want to know more about how well its graduates actually perform in their later careers. A campus or an entire system may be particularly committed to minority student success, so retention, achievement, and completion become a focus for assessment. Or legislators may be particularly concerned about the economic impact of an institution in a depressed area, and so outreach, retraining, continuing education, or the institution's role in attracting new businesses to the region may become a focus for assessment efforts.

Assessment can be undertaken at various levels; in other words, the "unit of analysis" can vary. We can, for example, undertake assessment in order to improve the performance of individual students; in that case, the focus is on each and every individual, with the line of feedback running from faculty member to student. More often, the intent is to use assessment to strengthen the performance of a program or service on campus, such as general education or academic advising. In that case, it may be necessary to look at only a sample of students; the feedback then runs to the faculty members or professionals providing the program or service. Students still benefit, but indirectly: not through direct feedback but through program improvement. Classroom assessment takes the class as the unit of analysis and attempts to improve the quality of teaching and learning in an individual course. At the institutional level, assessment looks at aggregate effectiveness and may include not just individual programs but the sum of curricular and extracurricular opportunities on campus, and consider the quality of administrative support as well as faculty and staff contributions.

Where and How Does Assessment Take Place?

There are various venues for assessment. For example, it may be embedded in course work: in assignments and in tests and quizzes. Students' work may then be reviewed by the instructor of record for content mastery, then passed on to secondary readers who look at it for evidence, for example, of writing or critical thinking skills. Or assessments may be administered to classes during class time but not necessarily as a part of the course; they are rather a time-out, then, from course work. If assessments are embedded in course work, the student's motivation to

work hard and do well is built in, because the student cares about his or her grade. The type of assessment that merely uses class cohorts in class meetings is liable to be less effective. Least of all are students motivated to show up and put out a good quality of effort when the assessment is an add-on requiring extra time in an unfamiliar setting, e.g., the gymnasium on Saturday morning. Assessments may take place off campus, particularly if we are looking at individuals whose lives are not centered on the campus such as alums, employers or parents of alums, commuting students, or those who have dropped or stopped out. There is no simple rule for what works and what does not, but in general, activities that are engaging, educational, and provide meaningful feedback to students (or others) work better than those that are routine, impersonal, and give little in return for the participants' donation of time and effort.

There are many methods of gathering evidence about the teaching and learning process, and assessment instruments and strategies proliferate by the day. To make some sense of this wealth of activity, it may be helpful to think about them in terms of purpose. In order to improve the educational process, we want to know essentially three things—though the following set of distinctions is a little artificial and in reality there is overlap among purposes and methods.

First, where are students when they come to us, either as entering students or as juniors or seniors entering a program of study or an advanced course? If we know something about students' academic history, we can support them in the areas where remediation or special attention is needed and provide appropriate and challenging course work beyond that point. It can also help us to document the amount of progress they have made over time with us, how much they have grown intellectually, or how much "talent development," in Alexander Astin's phrase, we have helped foster. Most institutions gather an immense amount of data in their admissions, registrar, and institutional research offices about the students they admit and how they progress through the institutions once there: SAT scores, socio-economic status, income, personal goals, attendance patterns, courses taken, grade point averages, and much more. Typically, this existing data is neither pooled nor exploited by institutions, yet it can be helpful in establishing base lines and setting priorities for further assessment. Institutions frequently begin their assessment efforts with a census of the kinds of information they are already gathering in various offices.

Second, we want to find out what students have accomplished during their time with us: we want to know what they know and what they can do with what they know; we may also want to know how their values have changed and what their habitual intellectual and affective responses are. There are many ways to gather evidence of student learning and development, and while tests or questionnaires are one way, there are also many alternatives. Current assessment practice ranges from standardized, commercially available instruments to locally developed tests and traditional essay-style comprehensive examinations. Samples of student work may be gathered and examined for evidence of various kinds of learning. Portfolios allow us to collect that student work for various purposes, for example, not only to demonstrate achievement but

also growth or progress over time; or to demonstrate not merely generic "writing skill" but skill in a variety of genres. Beyond that, assessment extends to performances in simulated settings (the assessment center approach), or real settings (the clinic, the machine shop floor, the classroom), or in internships and fieldwork. Certain kinds of courses—seminars, guided independent research, or capstone courses—can also serve as assessments. In capstone courses, for example, students pull together what they've learned and produce a final project. Such a course provides an opportunity for students to synthesize and demonstrate what they're capable of and, at the same, time allows faculty to think about the strengths and weaknesses of the preparation they have provided, based on the quality of the student projects. If external judges or examiners are involved in the evaluation of students' projects, they can provide a reality check for faculty as well as students.

Finally, we need to understand something about how students got to where they are. What are their perceptions of curricular and extracurricular life on campus? Do they understand why the faculty requires a particular kind of course or a particular sequence? Are students actually able to take courses in the sequence intended, and if able, do they do so? What elements of campus life do they experience as helping them or hindering them in their academic progress? To answer such questions, many information-gathering techniques have proven useful, from the traditional alumni/ae survey to wider-ranging surveys, interviews, focus groups, journal writing, and some classroom assessment techniques. Ethnographic-style, participant-observer research, in which students participate in instruction but also observe the classroom process and ask fellow students about their experiences, is an interesting new approach that has been applied chiefly to questions of retention. For example, why are talented potential science majors choosing other specializations? Why do Native American students drop out at a higher rate than others? Techniques adapted from Total Quality Management (TQM) are being introduced, both to help higher education better understand what its "customers" (students, employers, parents, and fellow faculty and other programs within the institution) really want, and as a rationale for gathering and analyzing data about the effectiveness of educational and support processes.

Assessment methods and approaches have proliferated in the last seven years in part because no single method can provide a full, multifaceted picture of the educational process in all its complexity. The corollary of that is that it's essential for assessment programs to make use of multiple methods in order to avoid reductionism and to capture accurately those complexities and nuances.

What Are the Current Trends in Assessment Method and Strategy?

There is a growing realization that quantitative approaches alone are inadequate and that qualitative evidence, in the form of narratives, descriptions, work samples, and demonstrations, must also be included

if we are to cover the entire range of educational outcomes that matter to us. We need to take care that we assess what is truly important, not just what is easy. Second, there is a growing trend toward customized, local approaches (as opposed to the use of instruments or models adopted wholesale from another setting) that truly reflect the curriculum, mission, goals, and culture of a given institution. Third, with the growing conviction that we must care not only about what students know but also about what they can *do* with what they know, there is a growing desire to look not only at students' paper-and-pencil evidence of mastery, but at actual student performances in naturalistic settings, whether on stage, in the clinic, at the front of a class, or on the job. Finally, there is a growing trend toward the embedding of assessment in everyday classroom practice, because it is less intrusive, cumbersome, and time consuming than add-on assessment, and at the same time more cost effective and motivating.

This brings assessment into the orbit of routine teaching and learning: "close to the classroom," as Pat Hutchings would say. In fact, I would claim that ultimately, if we carry this trend toward embedded assessment to its logical conclusion, virtually *any* classroom activity or assignment or student product can fulfill the requirements of a course and at the same time function as a valid assessment. The difference is what we *do* with that product or performance. Do we simply use student work as the basis for assigning the student a grade, or do we use it as the occasion for rich, personalized feedback to the student? Do we use student performances merely to judge the student's success in meeting our requirements, or do we use it as an occasion for critical, honest self-reflection and candid discussion of our *own* performance? Do we merely record and report a small slice of the student's educational experience, or do we actively try to put together a bigger picture of what all the pieces add up to, and then *use* what we've learned to develop more effective courses and programs? Newcomers to assessment often ask how assessment differs from ordinary grading, and this, in fact, is the answer: in assessment, we are looking at more than just an individual course, and we're looking at ourselves as well as students.

Education is a difficult, complex, and contradictory process. Any truly adequate assessment of educational outcomes must necessarily be difficult. It is relatively easy to assess students on their command of facts; it is much more difficult to figure out how we can assess the success of our educational program in developing students' critical abilities, their intellectual skills, their habits of mind and heart, the drives and emotions that motivate them to make difficult and unorthodox choices. And yet these are the very things that we may care most about. If we are going to use assessment to improve undergraduate education, the implication is clear. We cannot compromise our ambitions for education, including all those qualities that are so important but difficult to assess, to meet the level of existing assessment technology. On the contrary, we need to hold fast to our highest ambitions for education and work hard at refining our assessments so that, in time, they may begin to reach the level of our educational ambitions for our students and for ourselves.