Assessing College Student Learning

Evaluating Alternative Models, Using Multiple Methods

By Robert J. Sternberg, Jeremy Penn, and Christie Hawkins

With Case Studies by Sally Reed

Association of American Colleges and Universities

Assessing College Student Learning

Evaluating Alternative Models, Using Multiple Methods

By Robert J. Sternberg, Jeremy Penn, and Christie Hawkins

With Case Studies by Sally Reed

Association of American Colleges and Universities

Association of American Colleges and Universities

1818 R Street NW, Washington, DC 20009

ISBN 978-0-9827850-4-1

To order copies of this publication or to find out about other AAC&U publications, visit www.aacu.org, e-mail pub_desk@aacu.org, or call 202.387.3760.

Contents

Foreword

Virtually every college and university in the United States has a mission statement that describes the purposes and principles underlying the education provided to students who choose to attend. Most of these mission statements demonstrate significant commonality of purpose, but they also express the distinguishing characteristics of institutional identity. As a result, institutions adopt different mixes of programs, recruit different students, and pursue similar yet varied learning outcomes for their graduates.

In a 2009 survey conducted for the Association of American Colleges and Universities (AAC&U) by Hart Research Associates, chief academic officers reported that 80 percent of colleges have adopted broad learning outcomes that apply to all students. These outcomes include a wide array of cross-cutting skills and abilities valued by employers—writing, critical thinking, oral communication, quantitative reasoning, intercultural skills, information literacy, teamwork, civic engagement, integrative learning, and ethical reasoning. And, contrary to many recent public representations, seven in ten institutions reported that they assess these learning outcomes across the curriculum, primarily through the department or major; an additional two in ten reported developing cross-curricular assessment processes.

We know, in short, what we want students to learn and carry with them from their college studies. We also have begun to create, for the first time, a culture of inquiry about how well students are achieving essential learning outcomes across their studies. The challenge now is to make assessment an integral part of faculty and student work, and a significant resource in strengthening learning. At the same time, the research on cognitive development and deep learning is clear and overwhelming: if we want to help students retain and deepen their learning beyond the specific context of course events (e.g., the assignments and tests), then we need actively to engage them in their learning through more intentional educational practices. We typically learn better and retain more for longer periods of time when we engage in learning in direct ways, using multiple learning modes. Reading information is a beginning; reading and hearing about the information results in greater reten-

tion of information over a longer period of time; actively doing or making something with the information further increases retention; teaching others about the information and how to use it produces even deeper learning. This developmental progression has been described as engaging students in high-impact practices. Significantly, the emerging evidence on portfolios of student work suggests that applying knowledge, selecting examples or representations of students' own work, integrating learning from multiple sources, and reflecting on the process of learning, its quality, and the outcomes—the how and the why of learning—further strengthens student learning.

Mission statements, results of faculty and administrative surveys, and faculty research about learning all show that campuses—regardless of size, public or private ownership, or Carnegie classification—are focused on and committed to student learning. There is unanimity that the critical mission of higher education is to advance student learning. Why then is the assessment of learning so often divorced from or unrelated to the work that faculty do?

Just as students who are more engaged in their learning are more likely to experience enhanced, persistent learning, faculty are more likely to be engaged and invested in assessing learning when assessment processes provide information that improves their pedagogy and practice. If assessment processes do not provide faculty and students with useful information they can readily incorporate into their daily practices, higher education will continue to struggle with reporting results that have little relevance to the essential learning outcomes and performance levels expected of graduates as articulated by employers and faculty alike.

In the following discussion of the primary types of assessment used on campuses, Sternberg, Penn, and Hawkins emphasize the importance of assignments. This critical point about assessment, which is made in the context of a discussion of portfolios and authentic assessment, recognizes the fundamental centrality of what we ask students to do. Assignments, whether in the curriculum or the cocurriculum, generate evidence of demonstrated student learning. As Sternberg, Penn, and Hawkins argue, "if instructors do not give students opportunities to create materials that can go into the portfolios, student portfolios may look thin—not because the students lack value-added skills, but rather because they were not given opportunities to create products that would fit nicely into their portfolios." Assignments invite students to produce their best work in response to significant questions and information. If we don't ask students, if we don't assign students, to create their best representations of their learning, they will be much less likely to do so.

Assignments imply clarity about the learning students need to demonstrate, and they imply that we have expectations for what the outcomes of the learning will be. The role of expectations also is central to useful assessment. If we cannot clearly state or communicate

what students need to learn and at what level of competence, then we cannot be very intentional or precise in assessing the degree to which the learning has been achieved. By articulating expectations for learning, students and faculty now can have shared understanding of what the standards of accomplishment are.

Rubrics have been used for years to develop and articulate the key elements of learning outcomes at increasingly more complex and demanding levels of achievement. Rubrics exist for all of the essential learning outcomes that employers and faculty have identified as markers of a successful graduate (see www.aacu.org/value/rubrics). Initiatives such as writing across the curriculum, efforts to incorporate technology into information literacy, and AAC&U's own VALUE (Valid Assessment of Learning in Undergraduate Education) project have found high levels of agreement among faculty around the core elements or criteria of learning for a broad array of outcomes regardless of the type, size, or location of an institution. In addition, rubrics increasingly are being modified to incorporate newer technologies and media that go beyond print or written representation. These rubrics reflect the ability to capture student learning and to represent learning through multiple mediums spanning the diversity of modes of communication utilized by students and society today.

Rubrics exist for all of the essential learning outcomes that employers and faculty have identified as markers of a successful graduate.

The recent release of the Degree Qualification Profile (DQP) by the Lumina Foundation augurs a next step in articulating student achievement in higher education by presenting a framework for essential areas of learning and levels of student competence for associate's, baccalaureate, and master's degrees. The DQP takes the descriptions of performance criteria in rubrics and indicates the levels of competence or achievement that all students should have in order to receive a degree at the specified level. The DQP asks institutions to certify that students receiving a specific degree have demonstrated an expected level of accomplishment by the time they graduate. The DQP places assessment of learning across a broad set of essential learning outcomes at the center of faculty and student work, lifting up the centrality of linking our stated expectations for learning, the work students are asked to engage with, and the evidence of accomplishment at specified levels.

As Sternberg, Penn, and Hawkins argue, there is today a more robust set of approaches and resources available for assessing student achievement than existed even five years ago. As a result, institutions are struggling to determine the best approach for their respective circumstances. The first step in the decision process is to focus on what an institution

wishes to measure and why. Only after the desired learning outcomes have been identified can decisions be made regarding the tools for assessing learning. In selecting appropriate tools and methods for assessment, dollar costs are important, but it is equally important that the types of assessment data generated be useful both for faculty and for students in order to facilitate actions to enhance learning in areas of needed improvement and for addressing differential achievement among groups of students within an institution.

As the authors explain in part 1 below, different assessment tools measure different outcomes; no single assessment can, by itself, measure all that is important in higher education. For this reason, as the case studies presented in part 2 demonstrate, institutions that are taking the assessment of student learning outcomes seriously are using multiple measures in a variety of ways.

Today in higher education we have unprecedented agreement on the importance of improving student learning achievement, on the learning outcomes that are essential for success in a global economy and for a healthy civic life, and on the need to examine our practices to provide a high quality education for all our students. We have unprecedented research evidence and insight into the learning process, the relationships among pedagogies, intentional practices, purposes and expectations for learning, and the capabilities to capture and represent learning in ways that could not easily have been done in the past. As a result, the following examination of alternative methods of assessing college student learning not only challenges us, but it also helps us rethink what we are assessing, why we are assessing, and how we are assessing in order to achieve these important outcomes of a democratic society.

Carol Geary Schneider
President
Association of American Colleges and Universities

Terrel L. Rhodes
Vice President for Quality, Curriculum, and Assessment
Association of American Colleges and Universities

PART ONE

Evaluating Alternative Models

By Robert J. Sternberg, Jeremy Penn, and Christie Hawkins

On Alternative Models of Assessing College Student Learning:
Is There a "Best" Model?

"Tell me what you eat," Jean Anthelme Brillat-Savarin wrote in 1826, "and I will tell you what you are." Amid ever-increasing pressure for accountability, we might adapt this adage to higher education today by saying, "tell me what you assess, and I will tell you what you are." Assessment, curricula, and instruction are closely intertwined. As a result, our decisions about how to measure, report, and respond to data on student learning have substantial repercussions for how we teach, how we organize our classes, and how we are perceived by the broader public. Yet, too often we base our selection of assessment instruments and methods on pragmatic reasons of cost, reputation, ease of administration, and tradition, while ignoring critical issues of institutional purpose, mission, and identity. Both sets of considerations need to be taken into account.

We argue here that in measuring collegiate learning it is important to focus first and foremost on the specific qualities the college or university wishes to foster, and then secondarily on assessment instruments and methods—not the other way around. In other words, we believe it is essential to identify the learning outcomes that are valued by the institution *before* selecting a tool to assess learning. Students, their parents, and donors of scholarships spend enormous amounts of money on college education, and they are entitled to institutional accountability for learning outcomes. Yet, regrettably, assessment decisions are frequently made without serious consideration of what the various assessment methods actually measure. All assessments, even those presented as atheoretical, have an underlying theory of skills measured. It is important to bring these theories out into the open, rather than to leave them obscured.

The tendency to put the cart before the horse has plagued educational assessment since the early IQ tests (Sternberg 1990). As a result, many widely used tests do not measure all the characteristics their adopters may expect a learning assessment instrument to measure—the creative or practical thinking included in Robert Sternberg's (2003, 2010) theory of successful intelligence, for example, or the musical, interpersonal, or naturalist

intelligence included in Howard Gardner's (2006) theory of multiple intelligences. For assessment to be most effective, it needs to be based on a clear conception of what is to be assessed.

Our goal here is not to survey trends (see instead http://www.voluntarysystem.org), but rather to examine the available options for assessing student learning at the college level as well as the psychological theories of learning and achievement that underlie them. In other words, we seek to understand better what we are "eating" when we choose one of these options. Our examination is organized according to the assessment taxonomy developed by Maki (2004), which groups assessment methods into four categories: (1) standardized instruments and inventories; (2) indirect methods that focus on students' perceptions of learning and engagement; (3) authentic performance-based methods, such as portfolios; and (4) locally designed tests and inventories.

Standardized Instruments and Inventories

The first option is to use a standardized test, such as the Collegiate Learning Assessment (CLA), the Educational Testing Service Proficiency Profile (ETS-PP)—formerly known as the Measure of Academic Proficiency and Progress (MAPP)—or the Collegiate Assessment of Academic Proficiency (CAAP). Although there are many standardized tests available, we highlight these three because they have received much attention as a result of their inclusion in the Voluntary System of Accountability (VSA), "an initiative by public 4-year universities to supply clear, accessible, and comparable information on the undergraduate student experience to important constituencies through a common web report—the College Portrait" (Voluntary System of Accountability 2011). The CLA is intended to measure critical thinking skills in an authentic way: "To gauge summative performance authentically, the CLA presents realistic problems that require students to analyze complex materials and determine the relevance to the task and credibility. Students' written responses to the tasks are evaluated to assess their abilities to think critically, reason analytically, solve problems and communicate clearly and cogently" (Council for Aid to Education 2011). The ETS-PP measures skills in critical thinking, reading, writing, and mathematics in the context of the humanities, social sciences, and natural sciences; only critical thinking and written communication are included in the VSA College Portrait. The CAAP includes modules that measure reading, writing, mathematics, science, and critical thinking; only critical thinking and writing are included in the VSA College Portrait.

Tests such as the CLA, the ETS-PP, and CAAP offer several significant advantages. First, they have been normed on large and diverse samples of participants. Second, they are psychometrically sophisticated, and yet their results can be readily explained to laypersons with little or no background in psychometric assessment. Third, because the tests are standardized, it is possible to compare the performance of one's own students to that of other students; the tests provide a uniform basis of assessment for all students. And because the CLA also contains items that require the application of critical thinking skills to authentic problems, the items presented to students are somewhat more engaging than those in other, more typical standardized tests.

These three tests are similar to the ACT and the SAT, standardized tests commonly used for college admissions, insofar as they are based on a notion of intelligence as involving, largely, analytical or critical thinking applied to a knowledge base (Gardner 1983; Sternberg 1985, 2010). Intelligence tests measure verbal, quantitative, and spatial reasoning skills as well as one's level of general information. Although test publishers are not eager to make the association, the roots of these intelligence tests can be traced back to the Binet-Simon Intelligence Scales, which were first presented in 1905 as measures of the ability to succeed in school. ("SAT," once an acronym for "Scholastic Aptitude Test," has in recent years been used as an independent term with the letters not standing for anything.) Among some educators, the SAT and the ACT have acquired a dubious reputation, in part because they are (incorrectly) believed to measure fixed, innate traits, and in part because they show somewhat notable ethnic-group disparities—as do the other standardized tests considered in this section that measure roughly the same constructs.

Although the cosmetic appearance of this group of standardized tests has changed somewhat over the years, the constructs measured have not. The tests largely measure so-called "general ability," or *g*, and all tests of this kind correlate very highly with one another (Frey and Detterman 2004; Koenig, Frey, and Detterman 2008). For example, in one study, correlations of the CLA with the SAT had a median of 0.50 when the student was used as the unit of analysis and a total correlation of 0.88 when the college was used as the unit of analysis (Klein, Benjamin, Shavelson, and Bolus 2007). These correlations were not corrected for attenuation (unreliability), so the study underestimates so-called "true score" correlations. Although the CLA, the ETS-PP, and CAAP are similar in their surface structure to tests of critical thinking, such as the Watson-Glaser Critical Thinking Appraisal or the Cornell Test of Critical Thinking, the construct these tests actually measure is *g*.

Because tests such as the CLA, the ETS-PP, and CAAP are highly correlated with the SAT and the ACT (Klein, Liu, and Sconing 2009), one could consider, in group data,

covarying out (holding constant) the SAT or ACT scores. However, while solving one problem, this procedure creates another: Does the residual score, holding constant the SAT or ACT scores, still represent a construct of interest? And if so, what is this construct? The answers to these questions are not clear. And if the SAT and the ACT are only moderate indicators of learning, then the residual is likely to be, at best, only a weak indicator.

Data collected through the VSA show that the CLA, the ETS-PP, and CAAP are also very highly correlated with each other. So in the end, it does not matter greatly which test is used, as they all measure roughly the same thing: aspects of general intelligence. Where the choice of test may matter, however, is in comparing results from different colleges. The placement of a given college relative to others will depend on which schools used the same assessment.

Standardized tests are often pitched as measuring the "value added" by a college education, but whether they actually do this, and whether they do it in an appropriate way, is debatable. The question an administrator has to ask is whether this model of assessment is appropriate for college seniors who are going out into the world—perhaps to further schooling, perhaps to jobs. This model of measurement may be appropriate for assessing the analytical aspect of undergraduate education, but is it as appropriate for assessing other skills that may matter more later in life—expertise in one's major, creative skills, practical skills, wisdom-based skills, motivation, conscientiousness, or passion for a particular career?

Underlying all standardized tests is the assumption that the same learning outcomes are equally relevant for all students, regardless of what they have studied in college. The tests measure the same outcomes for engineers as for artists. For this reason, the abilities demonstrated by a highly successful music major, or engineering major, or even liberal arts student may not be reflected on a test like the CLA, ETS-PP, or CAAP. The risk is that, while this model of measurement might have been useful in predicting first-year grades in college from the SAT or ACT, the same model may be less useful later in assessing advanced learning through tests such as the CLA, the ETS-PP, and CAAP. For those students who go into careers where verbal reasoning is of only secondary or even tertiary importance, this model of measurement may seem to be lacking. An overemphasis on standardized measures like the CLA, the ETS-PP, and CAAP risks focusing our institutions on a narrow set of analytical and written communication skills that, while important, represent only a small subset of the skills and abilities we need to help our students develop in order to prepare them fully for later life.

Indirect Measures and Measures of Engagement

The second option for assessing student learning at the college level is to use an indirect measure, such as a student survey, focus group, or exit interview. One indirect measure that has received much attention in recent years—and is included in the VSA—is the National Survey of Student Engagement (NSSE), which is intended to assess the extent to which students are engaged in the college experience. Although "engagement" is not a well-defined term, it generally refers to the extent to which a student is involved with or even immersed in the college experience, including both academic and nonacademic components.

Although indirect measures are not explicitly theory-based, they seem to derive from motivational rather than ability-based theories. On this view, a student is successful in college to the extent that he or she is involved in the life of the college, both academically and nonacademically. Motivation is not totally unrelated to abilities, however. A typical Japanese view of ability is that it is, in fact, motivational at its core, and some psychologists studying intelligence, such as Carol Dweck (2007), have emphasized the importance of motivational elements underlying conceptions of intelligence, such as whether or not it is modifiable through hard work. In a similar vein, Cacioppo and Petty (1982) have emphasized the importance of "need for cognition," or the impetus to think and learn.

Although "engagement" is not a well-defined term, it generally refers to the extent to which a student is involved with or even immersed in the college experience, including both academic and nonacademic components.

There are volumes of data on NSSE, so it is possible to compare data from one's own college to data from other colleges. It is also possible to measure students' level of involvement in college. Many conventional types of measures are based on a model in which all that seems to matter are the academic aspects of a student's career. Academics are not the major focus of all students, however, and a measure based on a motivational model ought to recognize outstanding efforts in a wide variety of areas that may matter to college students.

NSSE is a typical rather than maximal performance measure (Sackett 2007), which presents a possible challenge to interpreting NSSE data. Ability tests are usually maximal performance measures: they measure how well one does when one gives the test one's all. In contrast, personality tests are usually typical performance measures: one indicates an answer, often on a Likert-type scale (as in a rating from 1 to 5). Such typical performance measures are "fakable" in that one can give an answer that one believes the examiner wants to hear, whether or not the answer truly characterizes one's own thoughts or behaviors.

NSSE is based on a model of measurement that values engagement, broadly defined, which presents a larger problem than faking. Students have very different career trajectories, both during and after college. One student may have been extremely engaged, but only in one activity—say, chorus. Another may have been extremely engaged in a project emanating from his or her studies. Yet another may have done lots of things, but not really have any core knowledge about an academic discipline. An overemphasis on indirect measures like NSSE risks rewarding students and institutions that are very broad but not necessarily very deep in their engagement, and it inadvertently can encourage manipulation of scores for the sake of appearances. In addition, indirect measures like NSSE do not directly measure any of the creative, critical, practical, or wisdom-based skills that are important outcomes of a college education. Instead, they provide only a glimpse of students' own perceptions of their level of performance— perceptions that may or may not be veridical.

Portfolios and Other Performance-Based Assessments

The third option for assessing student learning is to use portfolios or other performance-based assessments. Unlike the options described above, performance measures and portfolios—including electronic portfolios—do not measure the same outcomes for all students assessed. In a study at the University of Cincinnati, researchers found no correlation between CLA scores and critical thinking scores on portfolio measures (Eskoe, Hall, and Nicholas 2009). The idea here is that students have studied different things, have pursued individual trajectories, and hence have different strengths and weaknesses after their college studies. Each portfolio is thus an idiosyncratic product, allowing a student to highlight his or her strengths and, perhaps inadvertently, weaknesses.

Because portfolios allow students essentially to construct their own assessments, they measure motivational elements as well as many of the broad cognitive skills in Sternberg's (2010) and Gardner's (2006) cognitive theories. Portfolios thus combine aspects of the first and second options discussed above, and, theoretically speaking, they are more broadly based than either. For example, the theory of "successful intelligence" (Sternberg 1997) holds that people are successfully intelligent to the extent that they recognize and capitalize on strengths at the same time that they acknowledge and correct or compensate for weaknesses. Successful intelligence also requires creativity in generating new ideas, analytical ability in ascertaining whether the new ideas are good ones, practicality in implementing the ideas and persuading others of their value, and wisdom in ensuring that the ideas help achieve a common good over the short and long terms through the infusion of posi-

tive ethical values. Portfolio measurements have the potential to assess such a broad range of skills.

Although the products may be entirely different across students, it is still possible to create rubrics for scoring portfolios. The theory of successful intelligence would suggest that the portfolios be scored, at the very least, in terms of four attributes: creative thinking, analytical thinking, practical thinking, and wisdom-based thinking. Through its Valid Assessment of Learning in Undergraduate Education (VALUE) project, the Association of American Colleges and Universities (AAC&U) has provided fifteen rubrics that measure diverse aspects of performance, including ones that are similar or identical to what Sternberg has referred to as creative, analytical, practical, and wise thinking. For example, acquiring competencies, taking risks, solving problems, embracing contradictions, and thinking innovatively, as well as connecting, synthesizing, and transforming, can all be assessed using the VALUE rubric for creative thinking.

A major advantage of portfolios is that they allow students to highlight their own idiosyncratic strengths and to demonstrate their own understanding of what they have accomplished. A student who has done one or more major science or engineering projects, or who has composed a sonata, or who has written significant poetry or journalistic works will have an opportunity to demonstrate his or her achievements in a way that would not be possible through a conventional standardized test.

A major advantage of portfolios is that they allow students to highlight their own idiosyncratic strengths and to demonstrate their own understanding of what they have accomplished.

A drawback to portfolios is that they take a great deal of time to create, score, and code into a database. Moreover, some courses of study may lend themselves better to the assessments than others. A student majoring in English or philosophy may have a number of essays, papers, creative pieces, or other items that fit naturally into a portfolio. By contrast, a student majoring in engineering or agriculture may have many artifacts—such as models, seeds, or animals—that would be challenging to include in a paper-based or electronic portfolio. There may even be differences within a given broad field. Civil engineers may have designed bridges or tunnels, whereas chemical engineers may have comparable knowledge bases but fewer physical products to show for their studies. In brief, the measures may conflate opportunity to create products with quality of learning in the college years. Unless portfolios are carefully designed to incorporate a wide range of paper- and non-paper-based achievements, an overemphasis on portfolio measurement risks limiting the skills and abilities we emphasize in our institutions. Further, if instructors do not give students opportunities to create

materials that can go into the portfolios, student portfolios may look thin—not because the students lack value-added skills, but rather because they were not given opportunities to create products that would fit nicely into their portfolios.

Finally, portfolios can be very challenging to score, and scoring them requires considerable expertise. Although rubrics are available, it is not clear that they capture all or even most of the elements that make for a distinguished portfolio, especially because those elements may differ from one portfolio to another. Measurement will probably not be as reliable as that obtained from a standardized test such as CAAP or the CLA.

Locally Designed Tests and Inventories

The fourth option for assessing student learning is to use locally designed tests and inventories. Colleges and universities may choose not to use any of the standardized assessments and to create their own assessments instead. Such assessments may be constant across the entire institution, or they may vary from one department or program to another.

The most obvious advantage of locally designed assessments is their flexibility, which enables them to meet the perceived needs of the academic unit.

With local assessments, there are potentially as many underlying theories of skills as there are instructors who create the assessments. The strength in this is that the assessments are not narrowly circumscribed by a particular theory. The weakness is that it may not be clear what the assessments are actually measuring, at least in psychological terms.

The most obvious advantage of locally designed assessments is their flexibility, which enables them to meet the perceived needs of the academic unit. It is not clear that there is any one measurement that is relevant across all disciplines, and professors in different institutions have different goals for their students and serve different student populations. In one university, for example, the application of knowledge to the real world may be viewed as very important; in another, it may be viewed as strictly optional. Even within specific programs, faculty may have differing opinions on what is important to assess. Locally designed assessments can reflect all the different emphases of colleges, units, or programs, and if faculty members' diverse perspectives are incorporated during the process of creating the assessments, they can increase faculty engagement in assessment.

One significant disadvantage of these "homemade" assessments is their lack of standardization, which makes it difficult to compare results across academic units and to establish when "good enough" is "good enough." Another potential drawback is that academic units may not

enforce rigorous standards for the assessments or for their scoring, either because of a lack of expertise in measurement and assessment or because of a lack of resources fully to support this complex work. Strong leadership from the unit head is essential for ensuring the quality of locally developed assessments, but this too can be problematic if the unit head is uninterested or if there has been significant turnover in the unit-head position. While allowing for the incorporation of skills and abilities that are emphasized in each institution's mission, an overemphasis on locally developed measures risks creating the perception of sidestepping full accountability.

Which Assessment Method to Choose?

There is no single "best" approach to assessing student learning at the college level. All the methods considered here have strengths and weaknesses. If the goal is to measure an institution's success in educating college seniors, then, ideally, several different methods would be used in combination. However, it would be far more difficult to measure an individual student's learning or the "value added" by college education for an individual student; no currently available assessment method even comes close to achieving either of these goals. But ultimately, the first question to ask is not which measure to use, but rather what is to be measured. Since no college or university has the time or resources to measure all possible learning outcomes, a selection must be made.

If a measure based largely on somewhat narrowly based cognitive theory is required, then one of the standardized tests would be the best choice. If a motivationally based measure is needed, then a typical-performance assessment of engagement would be best. If a measure representing a mix of broad cognitive and motivational factors is appropriate, then portfolio assessments would be best. Locally designed assessments vary widely, and what they measure depends on how they are constructed.

At first, it may appear to be a sad state of affairs that the existing measures are so limited in what they can accomplish. But the outcomes of a college education are multifarious, and it is unlikely that there ever will be one perfect measure—because of the diversity of students as well as the diversity of skills, abilities, and disciplines. As is always true in measurement, using several different methods can compensate for the inadequacies of each. Using a variety of assessments almost always helps reduce error of measurement and, thus, enables focus on the constructs of greatest interest. An overemphasis on any one method risks distorting the assessment process itself, because no single measure can capture all the learning outcomes of a college education.

Unfortunately, college learning assessment has become a bit like ability assessment, with the cart coming before the horse. College administrators often pay more attention to the various measures than to the more fundamental question of what qualities they want to assess by using the measures. If there is a lesson to be learned from over a century of ability measurement, it is that we need first to decide what we want to measure, and only then to select among existing measures or create new ones. It is not very reassuring that some tests are now named by acronyms that stand for nothing; the danger is that no one knows quite what the tests measure—and worse, that no one cares. This is the legacy of a history of measurement that is unclear in its intellectual grounding. One may argue that the advantage of a test that is not explicitly based on a particular psychological theory of learning and achievement is that it does not favor any particular point of view, but in the end, tests always favor some point of view, whether or not their creators acknowledge it. Test developers may not specify the theory behind their tests, but implicitly, there always is one (or more).

Although cost is not inherent to any assessment method, which is why we do not discuss the issue above, any selection of measures will inevitably involve cost–benefit considerations. Portfolio measures are almost certainly the most expensive to design, administer, score, and process, simply because of the quantity of information they produce and the need to evaluate it subjectively. Locally designed measures are typically the least expensive, and the cost of standardized testing falls somewhere in between. The ratio of cost to benefit is something each institution has to compute for itself, depending in large part on how highly it values what is measured by a given assessment method.

In the absence of a perfect measure, some might wonder whether it is worthwhile even to assess college learning at all. We would argue that it most certainly is. It is true that no test given at any institution is ever perfect. Psychological and educational measures are, by their very nature, flawed and imprecise. However, college administrators owe it to themselves, and certainly to their students, to assess the value of the education they provide. Imperfect though they are, the existing methods described above are nonetheless useful ways of conducting such assessments. If you are what you eat, then colleges may indeed be—or at least be perceived to be—what they assess.

References

Cacioppo, John T., and Richard E. Petty. 1982. "The Need for Cognition." *Journal of Personality and Social Psychology* 42 (1): 116–131.

Council for Aid to Education. 2011. "CLA: Returning to Learning." Accessed May 23. http://www.collegiatelearningassessment.org.

Dweck, Carol. 2007. *Mindset: The New Psychology of Success.* New York: Ballantine.

Escoe, Gisela, Wayne E. Hall, and Mark C. Nicholas. 2009. "E-portfolios and Nationally Standardized Tests: Validation or Depreciation?" Presentation at the Sixth Annual Conference of the International Society for the Scholarship of Teaching and Learning, Bloomington, Indiana, October.

Frey, M. C., and D. K. Detterman. 2004. "Scholastic Assessment or g? The Relationship between the Scholastic Assessment Test and General Cognitive Ability." *Psychological Science* 15 (6): 373–78.

Gardner, Howard E. 1983. *Frames of Mind: The Theory of Multiple Intelligences.* New York: Basic Books.

———. 2006. *Multiple Intelligences: New Horizons in Theory and Practice.* New York: Basic Books.

Hart Research Associates. 2009. *Learning and Assessment: Trends in Undergraduate Education.* Washington, DC: Association of American Colleges and Universities.

Klein, Stephen, Ou Lydia Liu, and James Sconing. 2009. "Test Validity Study Report (TVS Report," http://www.voluntarysystem.org/docs/reports/TVSReport_Final.pdf.

Klein, Stephen, Roger Benjamin, Richard Shavelson, and Roger Bolus. 2007. *The Collegiate Learning Assessment: Facts and Fantasies.* New York: Council for Aid to Education.

Koenig, K. A., M. C. Frey, and D. K. Detterman. 2008. "ACT and General Cognitive Ability." *Intelligence* 36 (2): 153–60.

Maki, Peggy. 2004. *Assessing for Learning: Building a Sustainable Commitment across the Institution.* Sterling, VA: Stylus.

Sackett, Paul R. 2007. "Revisiting the Origins of the Typical-Maximum Performance Distinction." *Human Performance* 20 (3): 179–85.

Sternberg, Robert J. 1985. *Beyond IQ: A Triarchic Theory of Human Intelligence.* New York: Cambridge University Press.

———. 1990. *Metaphors of Mind: Conceptions of the Nature of Intelligence.* New York: Cambridge University Press.

———. 1997. *Successful Intelligence: How Practical and Creative Intelligence Determine Success in Life.* New York: Plume.

———. 2003. *Wisdom, Intelligence, and Creativity Synthesized.* New York: Cambridge University Press.

———. 2010. *College Admissions for the 21st Century.* Cambridge, MA: Harvard University Press.

Voluntary System of Accountability. 2011. "Voluntary System of Accountability Program: Undergraduate Education Reports." Accessed June 20. http://www.voluntarysystem.org/index.cfm.

PART TWO

Using Multiple Methods

By Sally Reed

Introduction:
The Cases for Assessment

In Florida, a college uses "real-life scenarios" as authentic assessments of students' communication skills or knowledge of ethical thinking.

In Minnesota, a small private college assesses the knowledge students gain in specific chemistry classes to measure learning outcomes and improve instruction.

In North Carolina, a university uses college-wide assessments as well as portfolios and other tools to address faculty concerns in academic departments.

Coast to coast, in small colleges and large universities, assessment is increasingly playing an important role in how students learn and teachers teach. Following on earlier publications and a series of articles in its member newsletter, *AAC&U News*, in spring 2011, the Association of American Colleges and Universities (AAC&U) decided to take a new snapshot of how different institutions have developed their assessment programs, including interviews with individual academic leaders guiding their campus assessment processes.

Case studies of five institutions with robust assessment programs are presented below. Representing a cross section of American higher education, they range from a private liberal arts college to a flagship state university. They address the issues surrounding the implementation of various assessment methods and the various challenges colleges and universities face. While hardly a comprehensive picture, these case studies provide a window into higher education practices in the current environment of increasing demands for accountability.

Background: Building on the LEAP Initiative

In 2005, AAC&U launched its major national initiative, Liberal Education and America's Promise (LEAP). LEAP's purpose is to highlight the importance of a liberal education, focusing on common learning outcomes for all students no matter their discipline of study or institution. The thinking was that as America entered a new century, it was important to

The Essential Learning Outcomes

Beginning in school, and continuing at successively higher levels across their college studies, students should prepare for twenty-first-century challenges by gaining:

Knowledge of Human Cultures and the Physical and Natural World

- Through study in the sciences and mathematics, social sciences, humanities, histories, languages, and the arts

Focused by engagement with big questions, both contemporary and enduring

Intellectual and Practical Skills, including

- Inquiry and analysis
- Critical and creative thinking
- Written and oral communication
- Quantitative literacy
- Information literacy
- Teamwork and problem solving

Practiced extensively, across the curriculum, in the context of progressively more challenging problems, projects, and standards for performance

Personal and Social Responsibility, including

- Civic knowledge and engagement—local and global
- Intercultural knowledge and competence
- Ethical reasoning and action
- Foundations and skills for lifelong learning

Anchored through active involvement with diverse communities and real-world challenges

Integrative and Applied Learning, including

- Synthesis and advanced accomplishment across general and specialized studies

Demonstrated through the application of knowledge, skills, and responsibilities to new settings and complex problems

---

Reprinted from Association of American Colleges and Universities, *College Learning for the New Global Century: A Report from the National Leadership Council for Liberal Education and America's Promise* (Washington, DC: Association of American Colleges and Universities, 2007), 12. This listing was developed through a multiyear dialogue with hundreds of colleges and universities about needed goals for student learning; analysis of a long series of recommendations and reports from the business community; and analysis of the accreditation requirements for engineering, business, nursing, and teacher education.

For more information, please visit www.aacu.org/leap.

articulate a broad vision for what college graduates need to know and be able to do to succeed in work, life, and citizenship.

A National Leadership Council, representing educational, business, community, and policy leaders and guiding the LEAP initiative, developed a set of essential learning outcomes as the core of this new vision. Listed here (see sidebar), the LEAP essential learning outcomes are a framework that articulates an overarching vision of the knowledge, the intellectual and practical skills, and the values and responsibilities that students need in a complex, global environment.

In commenting on the evolution of assessment practices in the California State University system, Ken O'Donnell, associate director for academic programs and policy in the Office of the Chancellor, noted that the LEAP essential learning outcomes "provide a common language not just within our system or community colleges, but across the country."

Indeed, over the last six years, the essential learning outcomes have played an influential role in the development of assessment practices at hundreds of colleges.

Conferences have focused on how to get started. Workshops have shared best practices. Publications and websites have described emerging challenges. Moving forward, rubrics, or the criteria used for assessing a skill or competency, have been developed through another AAC&U LEAP project called VALUE (Valid Assessment of Learning in Undergraduate Education).

But for many colleges, the work is just beginning. And to help them navigate their way, AAC&U has profiled five institutions to showcase the struggles and achievements of others who've already embarked on this journey and the different paths they have taken.

The Pressures and Passions

In none of these cases was the journey easy.

Sometimes colleges began by defining learning outcomes and assessment because of an interest among administrators or faculty. Some disciplines, such as engineering, education, and nursing, had licensing requirements or certifying agencies that dictated what should be included in their curricula. Some states articulated the need for greater assessment in colleges in response to the national movement to make higher education more accountable. But most often, college assessment programs got up and running because of a directive from one of the regional accrediting associations, the Southern Association of Colleges and Universities, for example, or the Western Association of Schools and Colleges. As O'Donnell noted, "the accreditors have played an important role in highlighting the value and indispensability of assessment."

But no matter the impetus, there were no roadmaps for these schools.

Often, creating assessment plans meant jumping over hurdles—budget restrictions, for example, resistance from faculty and staff, little administrative support, problems with matching instruments with the desired learning outcomes, sweeping changes in the disciplines themselves, and the difficulty of harnessing the new modes of the delivery of education. Each institution profiled here had problems along the way, but administrators and faculty worked them out within their own academic cultures.

No one way fitted all. Each institution met the need to measure outcomes in different ways. Some started by adopting the LEAP essential learning outcomes themselves. Others used them as a springboard to create their own set of common learning outcomes. Some used college-wide assessments as spot checks on specific aspects of their curriculum such as writing skills. Others began with assessing the knowledge gained in specific departments, such as psychology, or philosophy or they measured the competency of students in

a foreign language. For some, assessment became tied to an overhaul of a curriculum. And the instruments used to measure outcomes varied widely.

What the institutions profiled appeared to have in common though is a core belief that defining learning outcomes and measuring them is important to their overall mission. For some, it didn't come naturally. But once faculty and administrators were on board, once they saw that it improved teaching and learning, assessment often became their passion. Then those involved in the process conveyed their passion to others.

Collaboration is key. When assessment works, it is apparent that it does so because of the working relationships, and sometimes affection, of the key players.

But faculty buy-in is crucial. In many cases, faculty were resistant to engaging in new forms of assessment because they felt it an encroachment on their domain in the academy, believed it was simply judging them, and felt put upon for the time it took to make changes. Yet often, once they saw that assessment could make them better teachers and their students better learners, they became involved. Then many would reach into their own academic expertise to create an instrument or plan that best fit their institutions. And after initial steps were taken, more formal structures evolved.

Duke University in Durham, North Carolina, for example, now has a faculty assessment committee, a standing committee of the faculty governance body that meets with departments to assist them with assessment. Miami Dade College in Florida has faculty who now serve in specially created administrative positions charged with assessment. St. Olaf College in Northfield, Minnesota, has an assessment subcommittee of the faculty curriculum committee that makes ongoing recommendations.

Collaboration is key. When assessment works, it is apparent that it does so because of the working relationships, and sometimes affection, of the key players. Even more striking, often the people who got involved in the first place ended up being right for the job. Few of those interviewed here had experience in defining learning outcomes and designing assessment systems before they took on the task. But these individuals either came forth or, on hunches, were selected. And when the programs worked, their selection seems inspirational.

It also seems important that, by and large, the administrators charged with implementing assessment come from the faculty. They carry the historical threads and represent the soul of the institution, reflecting, rather than dictating, the sentiments of all those around them. They then have the intellectual authority to spread the mission.

Getting Started

Individuals quoted from the schools profiled here did offer suggestions about what they thought would be helpful to others just getting started.

Matt Serra, director of the Office of Assessment for Trinity College at Duke University, suggested that colleges starting out turn to the expertise in those colleges and departments that already have assessment experience with their accrediting bodies such as engineering, nursing, or education. "If any university or any college has an education program, they should take advantage of any expertise there," he recommended. "The same if you have an engineering school. And the same is the case with any medical school or nursing school. Nursing was way ahead of almost everybody with their accreditation and assessment."

When all is said and done, Pamela Menke, vice provost of Miami Dade College, argued that in addition to encouraging plans to grow organically from the faculty, it is beneficial to have meetings "well facilitated" by outside consultants who can help faculty work through tough issues, think of the institution as a whole, define terms and learning outcomes, and then form a consensus.

It is a lot of work. And Maureen "Mo" Bischof, assistant vice provost of the University of Wisconsin–Madison, observed that both in the past, and all too often today, it is an associate dean at a college or university who is asked to take on assessment planning on top of all his or her other duties such as academic advising or curriculum design. She suggested it is important that there be an "assessment person" whose job it is "to make connections and coordinate the various pieces."

None of those interviewed felt their jobs were finished. Today, assessment of learning outcomes is not a destination reached, but an ongoing journey. To paraphrase Lenore P. Rodicio, dean of academic affairs at Miami Dade College, the issue is not going to go away.

What follows are stories of how five very different colleges and universities developed assessment programs in their schools. They include: California State University–Sacramento; Duke University in Durham, North Carolina; Miami Dade College in Florida; St. Olaf College in Northfield, Minnesota; and the University of Wisconsin–Madison.

California State University–Sacramento:
Program Reviews Lead to Change

To Terry Underwood, the Program Review Oversight Committee at Sacramento State is at the heart of what is best about assessment at the university. It has had an impact on academic departments as they conduct self-study and reflect on assessments, said Underwood, who for the last three years has been the university faculty assessment coordinator. In fact, Sacramento State's mission is to build a program of assessment to enhance student learning and encourage excellence in teaching. But it wasn't always that way.

California State University–Sacramento, known as Sacramento State, is home to a multicultural student body. Seven percent of students are African American, 20 percent Asian/Pacific, 16 percent Latino/Latina, 1 percent American Indian, and 43 percent Caucasian. Thirty-three percent of the students consider themselves "multi-ethnic." A third of all students receive some form of financial aid. Undergraduates make up 77 percent of the nearly twenty-nine thousand students who study in sixty undergraduate, forty graduate, and two doctoral programs. Sixty percent of students in the Cal State system are transfer students. Sacramento State was founded in 1947 as the state capital's university, but in one generation it has grown in size and complexity and become part of a much larger statewide system.

The Backdrop of the CSU System

Thirty years ago, each university in the California State University (CSU) system set its own curriculum. But in 1980, a general education breadth curriculum was adopted for all campuses. According to Ken O'Donnell, associate dean for academic programs and policy in the Office of the Chancellor, the system had "a pretty decent tradition" of assessment.

"But as the culture in higher education shifted in this direction in the 1980s and 1990s, a couple of campuses were at the national forefront in this kind of work, such as Cal State Monterrey," he said. "It was built from the ground up around the idea that curriculum should be organized around student proficiency. The best of their work was to look at stu-

dent work and find ways to rigorously evaluate what the work showed in the way of student learning. They were good at developing rubrics and benchmark standards for performance. By their own admission, they had a harder time feeding the results of that assessment back into their work developing curriculum and pedagogy. So we had a good sense of what students were learning, but we didn't always use it to improve what we did."

The other problem was that as CSU campuses developed a curriculum built around proficiency, departments created different names for courses. On some campuses, "there was no psychology," said O'Donnell, "but Understanding Human Cognition. This was a disadvantage to transfer students. It wasn't clear how different courses should fit together."

O'Donnell noted that the different campuses shared best practices with one another. But more significantly, the accrediting agencies began putting more pressure on universities to formalize assessment programs. In 1996, Sacramento State started granting release time to a faculty member to do assessment work. "There was increasing external pressure to do more than just write assessment plans and put up wallpaper to actually doing assessment," Underwood noted.

In fact, the Western Association of Schools and Colleges (WASC) had been applying pressure to stabilize and make assessment meaningful for some time. Nonetheless, Sacramento State was reaccredited in 1999. "It looked like we were on the right road," Underwood said. "We had baccalaureate learning goals and assessment plans galore."

By 2004, a faculty member was released half time to be an assessment coordinator. However, "that person did not have an office," Underwood added, "but was sort of a wandering minstrel."

Underwood believes that when WASC returned for follow-up visits, they'd expected to see really good things. "What they saw was that there wasn't much going on at all," he said. "When WASC came in 2007-2008, most of the assessment plans on record had been written in 1999. There wasn't much evidence that anyone had actually implemented those plans."

Then, in 2008, the CSU enacted an executive order mandating that the entire system formally adopt, as the basis of the general education curriculum, the essential learning outcomes developed by the Association of American Colleges and Universities (AAC&U) through the Liberal Education and America's Promise (LEAP) initiative. In addition, the executive order said that campuses would assess the general education requirements.

To get a head start, many campuses also used the AAC&U rubrics developed for the VALUE project, which stands for Valid Assessment of Learning in Undergraduate Education. "Some used the executive order and the LEAP essential learning outcomes as a point

of departure and customized it in terms of its own faculty strengths, what students needed, and the requirements of the regional workforce," said O'Donnell.

An Office of Academic Program Assessment

Back in Sacramento that year, Terry Underwood was in the Department of Teacher Education and was granted increased release time from his classes to become the university faculty assessment coordinator for the Office of Academic Program Assessment. He had a background in assessment, a doctoral thesis on portfolio assessment, and had published on the topic. The university also gave him an office, photo copier, computers, telephones, and a graduate assistant. "We looked almost like a real operation," he quipped.

When WASC visited in 2009, Sacramento State received full accreditation. WASC saw that the institution had established an office and provided consultation. By this time, baccalaureate learning goals had LEAP outcomes embedded in them. Departments submitting reports had increased from 20 percent in the mid-2000s to pretty close to 100 percent. "They saw that we were on a trajectory," said Underwood.

Sacramento State is now fully accredited until 2018. But it is required to submit an interim report in 2012 covering four areas: the link between the analysis of evidence to planning and budgeting, the culture of advanced scholarship and assessment in a new doctoral program in education, how to address the problems of completion and retention rates, and assessment of general education. "The pressure is back on," said Underwood.

"We also had made changes in our Program Review Oversight Committee (PROC), a senate subcommittee in charge of program reviews," he added. "Those occur every six years for each department."

In fact, every six years, each department does a self-study. But now external reviewers are brought in. Then PROC writes a report with recommendations to the provost for each program. To Underwood, the Program Review Oversight Committee is key to the university's assessment. "Always before, those program review activities were dumped in the lap of the department chair," he said. "The faculty may not have even known it was going on. It called for a list of 140 little individual items that they had to report on. In the new process, there are three parts to it."

The first is a report on courses, student completion, number of faculty, and research. The second part is an analysis of assessment work over the last five years. In the new process, departments have to synthesize and reflect on what they have done in terms of assessment. For example, in the foreign language department over the past five years, the French faculty

did assessments in terms of conversation—how well students can speak French in a variety of settings. Faculty developed a rubric, conducted conversations with students, and examined student work.

In Spanish, the department examined student writing. "A lot of students speak Spanish in the home," said Underwood. "When they come to Sac State, they are fluent, but they can't write very well." The Spanish faculty focused on writing and then used assessments sustained over several years. "They made changes in their program because of that," he said. "That would not have happened had the change in program review policy not happened."

A "focused inquiry" is the third approach. A department engages in a large question it has a genuine interest in examining. For example, the Department of Family and Consumer Sciences had three concentrations: apparel and marketing, nutrition and dietetics, and family studies. "We're all under one umbrella," said Dianne Hyson, chair, "and that makes it really challenging to pull an assessment team together. We wanted things more integrated."

So Hyson worked with Underwood and nine faculty members to adapt a rubric that they thought would help them figure out how their program could be more integrated. They pulled together assessments from the different areas to study how to foster integrative thinking. "Assessment is in a whole different realm from when I started," said Hyson. "I think it is meaningful."

Revamping General Education and other Activities

Marching alongside the program reviews has been the revamping of general education at Sacramento State. According to O'Donnell, students fulfill the majority of their general education requirements in the first two years of a baccalaureate program. For many students, that means that the bulk of their credits are fulfilled at a community college. But the remaining credits must be fulfilled on one of the university campuses. Some students gather credits at several colleges before attending a Cal State campus. O'Donnell theorizes that one reason students may drop out of college is that they don't see the connections between the different requirements.

"Whenever we try to change education," said O'Donnell, "we have to take into account student community college mobility. It is significant. Assessing general education learning outcomes is critical. We want to make sure that our transfer curriculum is working because it is the only way of standing behind a degree."

Using funding from AAC&U's Give Students a Compass project, Sacramento State has spent the last few years assessing how to make general education work more efficiently and

how to redesign its curriculum. Sacramento State will have a new "GE pathway" to provide a "coherent program with integrated high-impact practices" beginning fall 2011.

"Now when we talk about learning outcomes assessment in the CSU, we try to build in questions of student mobility from the outset," said O'Donnell. In fact, he says that that is one of the appeals of the LEAP essential learning outcomes. "They provide a common language not just within our system or community colleges," he said, "but across the country, which we can use to map our traditional coursework so that we are getting an outcome focus without compromising transferability."

Conversations about assessment at Sacramento State led Underwood and others to examine areas beyond the essential learning outcomes in need of assessment, such as reading and math. Reading is of particular interest to Underwood, who is stepping down from his assessment position in fall 2011 to return to teacher education. He notes that the essential learning outcomes do not have a rubric for reading and would like to get faculty thinking about how to assess prior knowledge and higher-order thinking skills. "Most professors view reading as a very mechanical process," he said. "I don't think they have a feel for the complexity of the reading process."

So he set up a faculty inquiry group including professors from biology, religion, French, and education who met over a year and examined classroom-based research into student reading practices. The group drafted a rubric. "Now we are testing the reliability of that rubric," said Underwood.

Underwood also noted that one problem for transfer students is the math requirements. Sacramento State is also looking at ways to ensure that students develop quantitative reasoning skills in community colleges. The university received a second Compass grant to work with a community college to develop an electronic portfolio platform that will start in a freshman seminar and move with students as they transfer to Sacramento State. They will then enter an academic learning collaborative, be part of a cohort, and continue to build a portfolio.

Sacramento State is indeed on a "trajectory," but administrators on its campus and in the chancellor's office are worried about the impact that state budget issues are having on assessment. "Most of the folks in the CSU who give assessment serious thought would say it is good to track student learning whether or not it's because accreditors insist we have to," said O'Donnell. "But so many of us are overworked, time is tight, and doing something because it is good doesn't get you very far. The accreditors have played an important role in highlighting the value and indispensability of assessment. We wouldn't have gotten far without them."

Duke University:
Creating a Culture of Evidence

The biology department at Duke University's Trinity College had a problem. The writing produced by undergraduate students for the capstone honors thesis was "hit or miss," as Julie A. Reynolds, associate director of undergraduate studies, described it. So armed with a PhD in biology and post-docs in writing, she devised a Biology Thesis Assessment Protocol for student writing dubbed BioTAP. It has proven so reliable she is now using it in the economics and chemistry departments.

Duke, located in Durham, North Carolina, is a highly selective private research university with an undergraduate enrollment of 6,504 students in two schools: Trinity College of Arts and Sciences and Pratt School of Engineering. Eighty percent of undergraduates are at Trinity; about half of its undergraduates receive financial aid. Duke also supports ten graduate and professional schools and seven research institutes. The biology department alone has about fifty-five faculty members. Yet, when Reynolds wanted to design an assessment tool, she turned to Matt Serra, director of the Office of Assessment, whose mantra in working with faculty is "to remove the roadblocks." In fact, he's been removing roadblocks for over a decade.

Creating an Office of Assessment

Matt Serra was a professor in Trinity's Department of Psychology and Neurosciences in 2000, when Duke revised its entire curriculum. "Then Dean Robert Thompson asked me if I would like to start an office of assessment," Serra remembered. "I said, 'Well, I don't really know what assessment is.' He answered, 'Well, I don't either, but we are going to figure it out together.' So we opened an office of assessment in conjunction with the new curriculum in order to assess the veracity of that curriculum."

Curriculum 2000 is the College of Arts and Sciences' inquiry-based, interdisciplinary curriculum, which has writing and research as its cornerstones. Its mission: to "establish an

undergraduate culture of research to directly link undergraduate education to the processes of inquiry and scholarship that characterize a research university." It further seeks to "establish a culture of evidence that reflects experimentation and an iterative approach to continuous quality enhancement through evaluation of learning outcomes."

To define those learning outcomes, Serra began his new job by looking at what the committees charged with developing this new curriculum determined was important. "We wanted to align the curriculum with the outcomes wanted," explained Serra.

It took several years to get the curriculum in place and to get learning outcomes developed in concert with the curriculum. But "at the time, it was totally unique," noted Serra, "in that every course in the Duke curriculum can count both as a general education requirement and as a course in an individual major, minor, or certificate program. It is all integrated."

The student learning outcomes for general education, for example, include critical and analytical thinking, ethical and moral engagement, civic engagement, global citizenship, foreign language, and quantitative reasoning, according to Serra. The Association of American Colleges and Universities' Liberal Education and America's Promise (LEAP) initiative helped provide guidance over time in terms of the detail of what undergraduates should know. Then there is ongoing assessment and "the curriculum is adjusted as we go," Serra explained.

The biology department added learning outcomes to address its own needs. Faculty focus groups defined what they wanted students to know and do. What they wanted was a more systematic way to talk with students about writing. Reynolds worked with Serra to create rubrics for this goal and align it with university learning outcomes. "Matt helped me think through the different categories," said Reynolds, "and not use language that is too vague. What's important is that different readers can pick up this rubric and use it consistently. So we have to have language that's not too ambiguous."

Creating a Culture of Evidence

The Office of Assessment next created a "culture of evidence" by developing a detailed assessment plan and has stayed close to it. The plan tracks indicators and benchmarks for all students. A database houses student characteristics, experiences, and academic performances. "We knew it had to be systematic," said Serra. "We took off a chunk at a time, something that could build on other pieces."

For example, Duke has been assessing foreign language for the last six years. "On the back burner, we always have going things such as ethical and moral development assessment," said Serra. "Now we are about to move on to quantitative literacy."

In fact, for the last year, a faculty committee has been reviewing the quantitative literacy outcome, which is an ability to think through not just data but quantitative information—poll numbers, for example, or information on camping water purification. The Office of Assessment found out how faculty and students go about fulfilling that requirement. "We don't expect everyone to be a PhD in calculus," said Serra. "But we expect everyone to function as a useful person in our society."

In addition, the psychology department is giving a test first to introductory-course students and then graduating seniors to see how they do over time. The philosophy department is giving "a knowledge exam" to graduating seniors. "The beauty of it at Duke is that the programs do their own assessment," said Serra.

In biology, BioTAP includes a rubric assessing thirteen elements and a guide for interacting with students. Student results are assessed as "not having satisfied" a learning outcome, "met minimum departmental standards," or "achieved mastery." The rubric helps "everybody go through the process more efficiently," said Reynolds, who is also an assistant professor. "Faculty have a common language. We can all talk with students about their writing."

College-wide, students fill out Student Course Evaluations, which are available to faculty. Trinity has administered the Defining Issues Test (DIT-2), which assesses student moral development; the Reasoning about Current Issues (RCI) test, which evaluates student reflective judgment; and the Collegiate Learning Assessment (CLA), which measures the development of student critical thinking skills.

"Every course in the Duke curriculum can count both as a general education requirement and as a course in an individual major, minor, or certificate program. It is all integrated."

Using Multiple Methods

According to Serra, the assessment tools have to be multi-methods. "I do what I call 'triangulations.' I bring two or three measures to bear on a specific issue. If they point in the same direction, fine. But if they don't, I figure out why. We use direct and indirect measures to get at everything. We don't rely on either one or the other. And we don't try everything every year."

Duke uses oral proficiency interviews to get at the foreign language attainment and standardized exams such the Global Perspectives Inventory to compare Duke students with others nationally and internationally. Duke does its own writing assessment. It uses electronic portfolios and collects artifacts including essays, blogs, and works of art or music. "We can collect

anything the faculty and students feel demonstrate they have attained an outcome and develop rubrics," said Serra.

The Office of Assessment not only consults with departments such as biology, it helps faculty and staff develop surveys or do data analysis. For example, Amy Anderson was hired two years ago as an instructor in the Program in Education and faculty consultant for the service-learning program. Her charge: to help develop assessment for Duke's service-learning program involving about forty-five faculty members in twenty disciplines throughout the university. She said that Matt Serra "was one of the first connections I made when I came to Duke."

Serra collected end-of-course evaluations and provided Anderson an analysis of the past five years on how students rated service-learning courses in comparison to other courses. The evaluation data also included "snapshots of who service-learning students are," she said. "And Matt was a sounding board to me about what I am trying to do. He also participates in our working groups."

A student advisory group told her that students didn't always know what they were getting into in service-learning classes based on the descriptions in online registration.

Anderson now does pre-and-post surveys of students in service-learning courses and examines students' reflections on their community work through their writing, verbal reflection, peer reflections, blogs, or performance-based instruments. A student advisory group told her that students didn't always know what they were getting into in service-learning classes based on the descriptions in online registration. "So we surveyed students," Anderson said, and now she is designing training for service-learning faculty on how to be more specific in the descriptions of courses. A faculty conversation group meets monthly for professional development and readings. "As we learn more from students, we try to bring those things to our conversation," she said.

Serra noted that faculty will have excuses for why not to do assessment—"'I don't have time,' 'I don't have the resources.' Well, we've got the resources and we can help alleviate the time demand. I tell people, 'You do the heavy intellectual lifting and we will help you with the other 'stuff.'"

Assessment Tied to the Budget

Finally, each department must produce an annual assessment plan and a detailed assessment report, which are tied to the annual budgeting process. And as part of Duke's last accreditation review, "we were required by our accrediting body that each of our academic depart-

ments and programs within Arts and Sciences have rigorous assessment of the outcomes of their majors," explained David Malone, associate professor, Program in Education. Malone was named chair of the faculty assessment committee, a standing committee of the faculty governance body. Matt Serra also sits on the committee.

The committee developed a plan for infusing assessment into Arts and Sciences and then a rubric for evaluating and reviewing the strategic assessment plans and reports that the departments were asked to do. In spring 2011, eleven faculty members began meeting with the forty-seven departments to "create a greater sense of importance," said Malone. "We view our mission as educational in terms of meeting with academic departments and talking about the benefits of assessment to them…. And we also see our mission as reviewing the work that they are required to write and providing feedback."

Malone stated that the committee focuses on four areas: Has the department articulated their learning outcomes? What direct and indirect measures do they use? Are the findings shared with others in the department? And have they used the findings for program modification to inform current practices?

While there is a dictum that departments do assessment, Serra notes there are other motivating external and internal factors. The Southern Association of Colleges and Universities continues to apply pressure to make program assessment even more systematic. There are pressures from the public, the government, and from granting organizations. "But internally, a professor wants to be a better instructor and improve pedagogy," said Serra. "You need the right combination. External motivation combined with internal motivation is key. Then we try to make the road as smooth as possible."

In the end, assessment to Julie Reynolds "is about the scholarship of teaching and learning. How do you know if what you are doing in the classroom is effective? You have to be rigorous about stating what you want to accomplish and then testing your assumptions."

Miami Dade College:
Assessment Supports the American Dream

In spring 2011, Miami Dade College (MDC) made national news with the announcement of its "American Dream Scholarship," a financial aid program that provides two years of tuition-free education for any high school student in Miami-Dade County, Florida, who graduates with a "B" or better grade point average and who passes an entry skills test. While the scholarship paves a new path of opportunity for students, it also establishes clear standards for them. The bold initiative is only the latest at MDC to expand learning options while setting academic goals that the college then assesses.

Miami Dade College is one of the largest higher education institutions in the country, serving 174,000 traditional and nontraditional students on ten campuses. It offers over three hundred two- and four-year degree, certificate, and nondegree programs. The majority of students have Hispanic backgrounds, and many students come from low-income and/or immigrant families. Founded in 1960, MDC's mission has been "to change lives through the opportunity of education." And its president, Eduardo J. Padrón, has gained an international reputation for his unrelenting pursuit to make MDC "accessible" and renowned as a "learning college."

Goals Led to Outcomes

This desire to create "a learning college" led MDC down its own path of rethinking goals and later assessment. "In 2005, the college got very interested in the concept of a 'learning college,'" said Pamela Menke, vice provost. In exploring that concept, it held symposiums with speakers such as Lee Shulman, then president of the Carnegie Foundation for the Advancement of Teaching.

"At the same time, MDC had become interested in AAC&U's liberal learning initiatives," added Menke. "Between the liberal learning activities and the desire to be a 'learning college,' we began to rethink what we were doing."

Against this backdrop, the Southern Association of Colleges and Schools (SACS) asked MDC to provide evidence that its graduates were attaining competencies in its general

education program. MDC's initial response was to dust off its old set of goals—some stemming from the 1970s, most established in the 1990s. By 2005, the accumulated list of random goals had grown to twenty-five. The problem was no one was aware of their existence.

"As a faculty member, I didn't even know we had general education goals," said Lenore P. Rodicio, a chemistry professor turned dean of academic affairs. "They were listed in our catalogues, but unless you were here at the college when they were developed, many of the faculty didn't even realize there was this list. Then when you read them, they looked outdated. One goal read, 'Students will use their leisure time wisely.' We asked, 'How are we going to assess these?'"

Rodicio said the faculty decided it needed more training and a look at what other institutions were doing. Twelve administrators and faculty, including Rodicio, attended two AAC&U conferences that year. "The more we listened, the more we realized it wasn't that we needed to assess what we had better," added Rodicio, "we needed to start from scratch."

S. Sean Madison, an English professor and now director of learning outcomes assessment, noted, "We could have easily pulled something off the shelf, administered it, and been done with it. But it would not have encouraged the kind of engagement we felt we needed for faculty to take ownership of this process."

The team of twelve convinced other faculty members that MDC should revise what it was still calling "goals." So at the start of the academic year, the faculty was asked, "'What does the MDC student of the twenty-first-century look like?'" said Rodicio . "'What should their skills and abilities be?' We just planted that seed in everyone's mind."

MDC then embarked on a yearlong process of articulating its learning goals. In addition to students and teachers, the college asked local business leaders and alumni to describe the skills that MDC graduates should be able to demonstrate. The college created focus groups, conducted surveys, and organized summits. The training and development center brought in nationally known facilitators such as Peggy Maki who met with groups of faculty to discuss potential goals. By summer 2006, the new goals were whittled down to eleven. An academic leadership council, composed of deans, reached further consensus on ten.

Then, in fall 2006, something "magical" happened, according to Menke. While she was meeting with President Padrón about the council of deans' agreement, she said he noted that the word "goal" didn't work. They were "learning outcomes," he argued. And, he added, they were not just general education goals. They should be college-wide learning outcomes. "The whole concept shifted," said Menke. The outcomes "involved every part of MDC and it took like glue."

That fall, students, faculty, and staff participated in a ceremony to sign a "Covenant of Engagement" to support the university-wide outcomes. Attending the ceremony were a union president, the CEO of the Greater Miami Chamber of Commerce, and a US undersecretary of education. The college put up posters about the outcomes, inserted them into the school catalogue, and a learning outcomes coordinating council was established with about thirty faculty members to keep the initiative going.

"Enter SACS," said Menke. The association informed MDC that while the work on developing the learning outcomes was good, the college needed to show how the learning outcomes would be assessed. Menke noted that MDC was heading in that direction already, but the request "catapulted us forward."

Designing Authentic Assessments

Indeed, MDC put the task of creating assessments "on a fast track," said Menke. The college also made what turned out to be a critical decision: assessments would be college-wide first. Again, working in teams, the faculty collaborated on aligning assessment with MDC's new "Ten Essential Learning Outcomes" (see sidebar).

According to Joanne Bashford, associate provost for institutional effectiveness, the faculty immediately began creating "authentic assessment" activities because "authentic activities involve many skills and are more meaningful than standardized tests," she said. "By early spring 2007, we were actually doing pilot assessments."

One authentic assessment activity, still in use today she noted, asks students to focus on nearby Biscayne Bay, a national park. The scenario describes the ecology of the reefs and centers on the prospect of oil drilling in the bay. The prompts provide data on oil drilling, pro and con. Students are asked to analyze

Miami Dade College's Ten Essential Learning Outcomes

1. Communicate effectively, using listening, speaking, reading, and writing skills.
2. Use quantitative analytic skills to evaluate and process numerical data.
3. Solve problems using critical and creative thinking and scientific reasoning.
4. Formulate strategies to locate, evaluate, and apply information.
5. Demonstrate knowledge of diverse cultures, including global and historical perspectives.
6. Create strategies that can be used to fulfill personal, civic, and social responsibilities.
7. Demonstrate knowledge of ethical thinking and its application to issues in society.
8. Use computer and emerging technologies effectively.
9. Demonstrate an appreciation for aesthetics and creative activities.
10. Describe how natural systems function and recognize the impact of humans on the environment.

the situation and write an essay. They are assessed on their communication skills, computer literacy, knowledge of ethics and social responsibility, and of sustainability, all on the list of outcomes. Other authentic assessment activities may involve public speaking, the use of graphs, the interpretation of works of art, or an ethical issue faced in the workplace. "Those are the kinds of problems our faculty feels will be engaging enough for students to capture their interest," said Bashford, "so they will be motivated to perform well on the assessment. Those are also scenarios they might encounter somewhere in conversations, in listening to the news, or in a real life environment."

Other authentic assessment activities may involve public speaking, the use of graphs, the interpretation of works of art, or an ethical issue faced in the workplace.

Since 2007, MDC has targeted 10 percent of each graduating class for assessment activities. Most students, around ten thousand this year, come from the courses that 75 percent of students took to complete their degrees. The subject of each assessment is randomly assigned and does not correspond to the subject a faculty member who administers the test teaches. Students spend fifty minutes responding to what President Padrón has called a "challenging, multifaceted scenario" of "authentic tasks" that deal with the "real-world situations" and have "no perfunctory answer."

Student responses are scored as either "Emerging, Developing, Proficient, or Exemplary." Results are aggregated to provide a "snapshot" of each graduating class's level of accomplishment. This is then compared to other classes. There are formal campus dialogues each year to discuss the findings and see what changes need to be made. A learning outcomes assessment team was formed to direct assessment activities.

Tracking Progress with "A Curriculum Map"

To further track the progress on providing "Ten Essential Learning Outcomes," MDC "mapped its curriculum." Instructors can see how students are doing as they go through their program. To date, about three-quarters of the two-thousand-plus courses have been charted on an Excel spreadsheet with ten columns that correspond to the ten outcomes, according to Menke. When areas of weakness are identified, the faculty can make course-specific adjustments. Courses are scored on how well they introduce, reinforce, or emphasize specific outcomes.

Faculty members are not responsible for all ten learning outcomes in a course. John Frederick, a former professor of speech, noted that in the speech department, for example,

"even though there are ten outcomes, we felt that as a discipline we contributed to seven outcomes. For example, as students debate, they use 'critical thinking.' 'Aesthetic appreciation' is part of the speech-making process and the study of the beauty of language."

MDC is now moving beyond its college-wide assessments to creating assessments at the individual program level. The learning outcomes are used to evaluate courses or when adding new ones. The faculty have been urged to "weave the learning outcomes and relevant assessments" into their individual classes. In March 2011, a "Faculty Assessment Showcase" featured examples of "best practice." And Frederick was recently appointed an administrator to assist programs in developing assessment.

But the ultimate decision on how to better attain these outcomes remains in the hands of departments and individual course instructors. According to Madison, from the onset, MDC's administration wanted "to really make this a learning opportunity for faculty. This is not about being punitive or trying to identify a faculty member who is teaching poorly. It is really about the student. We established some guiding principles. It is all about learning improvement. Are we fulfilling our mission as a college?"

MDC's belief in assessment is now carved in stone. It is the second point in its eight-point vision statement, and the vision statement is included in the college's strategic plan. And its learning outcomes and assessment work have led to several accolades: the 2008 Exemplary Award for Improving General Education from the Association for General and Liberal Studies; the CollegeKeys Compact 2010 Innovation Award, selected by the College Board; and the 2011 Council for Higher Education Accreditation Award for Outstanding Institutional Practice in Student Learning Outcomes.

What does the future hold? According to Rodicio, "one key question is, 'how do we use the data to continually improve the classroom?' Also, 'How do we keep faculty going?' In the beginning, some faculty said, 'This is a fad, it will go away.' People have realized it is not. Assessment has become engrained in the fabric of the college."

St. Olaf College:
The Transformative Effect of Faculty-Driven Assessment

Last year, the Council for Higher Education Accreditation (CHEA) gave St. Olaf College in Northfield, Minnesota, its Award for Outstanding Institutional Practice in Student Learning Outcomes. It noted St. Olaf "uses evidence of student learning to inform faculty development, increase collaboration, and inspire instructional improvement."

Yet in 2003, when the Higher Learning Commission of the North Central Association of Colleges and Secondary Schools granted St. Olaf reaccreditation, it said that its assessment program "requires institutional attention." It further required St. Olaf to provide a written progress report on the program by January 2007.

How did St. Olaf take an assessment program from "requiring attention" to becoming a national model? To Jo Michelle Beld, the director of evaluation and assessment at St. Olaf, it's because the college's approach to assessment became "mission driven, meaningful, and manageable."

"Feet to the Fire"

St. Olaf College was founded in 1874 by Norwegian immigrants and is a private four-year liberal arts college of the Evangelical Lutheran Church of America. In fall 2010, it had 3,156 students, nearly half from Minnesota. The college offers 39 different disciplines and subject areas and 20 interdisciplinary concentrations. Students may study in 110 locations around the globe. It employs 213 full-time and 123 part-time faculty members and is internationally renowned for its music and science programs. Its mission is centered on "a commitment to the liberal arts, rooted in the Christian gospel, and fostering a global perspective."

St. Olaf began working on assessment in 1993. "Like many institutions, we started having our feet held to the fire in the early 1990s," said Beld, who was then a faculty

member in the political science department and later became director of general education. "When we had an accreditation review in 1993, we were asked to start working on assessment. Everybody was making plans and identifying learning outcomes, but no one really knew what they were doing. Most of us in the trenches thought it was done to make our accreditors happy. That is not a very inspiring way to look at assessment."

Then in 2003, with another reaccreditation review, "the bar was raised and we were pretty far behind," she said.

St. Olaf established an office of academic research and planning directed by Beld, who began gathering materials such as course evaluations and tenure reviews. Eric Lund, professor of religion and now director of international and off-campus studies, recalls that there'd always been some kind of program evaluation. But evaluation of study abroad programs, for example, just included questions about the logistics of the experience such as the housing accommodations. Then departments were asked to do a "self-study" and analyze why things hadn't gone well and what they hoped to do in the future. "That was a turning point for us," said Beld.

For Mary Walczak, professor of chemistry, the real turning point was when Beld "started using the phrase 'inquiry in support of student learning.' Who could argue with that?" she asked. She said that indeed the faculty had been distrustful about prior assessment initiatives. But once the faculty began to feel there was guidance on how to do assessment, "it changed the tide," she added. "The culture changed from assessment being the 'A' word to using the 'A' word. A critical mass of the faculty became supportive of the enterprise."

Beld was next named director of a new department of evaluation and assessment at the college, charged with overseeing course evaluations, but also the assessment program, student evaluations of faculty, academic program reviews, and grants for evaluation planning. The Teagle Foundation funded St. Olaf to create new assessment instruments in partnership with Carleton, Macalester, and Grinnell Colleges. "It was the jump-start we needed," said Beld.

The college was further funded by the National Institute for Technology and Liberal Education to examine ways to improve student literacy. Working with other colleges, St. Olaf developed a research practices survey to examine literacy at the beginning and end of a student's college career. It administered the Collegiate Learning Assessment (CLA) and the National Survey of Student Engagement (NSSE). By 2008, St. Olaf had an assessment plan. And Beld was released from teaching courses to manage the assessment work.

Getting the Faculty Engaged

"Two things then helped us moved forward," said Beld. "Having me as a faculty member lead the effort, and making it a mission to make the assessment useful to faculty."

Beld's hope was that the faculty would not only engage in assessment, but welcome it. She argued that the assessment had to be meaningful to not only improve student learning, but promote faculty development, satisfy the intellectual curiosity of those charged with delivering programs, and inform those who make decisions about programs, majors, or requirements. Moreover, it had to be manageable.

Yet as Beld started making requests for assessment from the Deans Council, she quickly saw that the requests needed to come from the faculty itself. Making assessment part of faculty governance turned out to be critical. St. Olaf established and now uses a structure whereby an assessment subcommittee of the faculty curriculum committee makes recommendations about the evidence that needs to be gathered. The curriculum committee gives these recommendations to the dean who puts out the requests.

"Each department defines what it does and what it expects students to do," noted Walczak. "Faculty members are given the freedom to define what learning is in that discipline. This has been key to reassuring faculty that it has been faculty driven."

A variety of other tools are used to analyze curriculum requirements, instructional strategies, programs of study, and services of the college.

It also became important that the department chairs and faculty not feel that instruments were constantly "thrown at them." So in 2008, when St. Olaf began implementing its plan, it also produced a schedule—running through spring 2015—showing what tests would be administered (and when) at the institutional level, by department, and in general education. Built into the schedule are times for reflection on the assessment results.

When St. Olaf first administered college-wide assessments, it used as its guide the essential learning outcomes developed by the Association of American Colleges and Universities (AAC&U) through the Liberal Education and America's Promise (LEAP) initiative. It evaluated critical thinking skills, for example, quantitative reasoning skills, and whether students were proficient in research. At least seven instruments have been used for college-wide assessment. Keeping to its schedule, St. Olaf administered the CLA in 2009, again in 2010, and will repeat it in 2013. It administered NSSE in spring 2011 and will repeat the survey in spring 2014.

A variety of other tools are used to analyze curriculum requirements, instructional strategies, programs of study, and services of the college. This evidence is gathered from

tests, portfolios, and interviews or from an analysis of the content in student papers. The assessment program also looks at general education requirements, such as first-year writing courses or foreign language requirements. And then the assessments are used to examine courses, departments, and programs.

In fact, this is now the heart of the St. Olaf assessment program.

Making Assessment Manageable at the Departmental Level

Beginning in 2008–09, and then again in 2009–10, individual majors, concentrations, and other academic programs were evaluated. Using the faculty governance structure, departments were asked about the majors they offered. In American Studies, for example, the department stated, in part, that its students were to demonstrate an understanding and use of terms and theories, engage in independent research, and be familiar with a range of disciplinary approaches.

To keep the process manageable, St. Olaf adopted "utilization-focused evaluation" developed by Michael Quinn Patton, an expert on organizational management and evaluation and author of a book of that title published by Sage Publications. According to Beld, "'utilization-focused assessment' turns on the core question, 'What evidence of student learning do we need to help us identify and sustain what works, and find and fix what doesn't?'"

In this model, the number of questions asked of a department is limited. Departments, in turn, are told to limit the number of outcomes they desire to five and the forms of evidence of student learning they will use to one. When departments submit reports, they report on what they are going to do with data, rather than data itself, and the data is in an appendix. In 2010–11, department chairs and program directors were spending the school year on "reflection, integration, and planning," examining the evidence gathered and then articulating what would they want to change.

"As a department, the psychology faculty decided they wanted to know what students were getting out of their majors," said Gary Muir, professor of psychology and neuroscience. "We want our students to have content knowledge in certain areas, and how do we know they are coming out with that?"

So three years ago, the psychology department began giving outgoing seniors a standardized test called ACAT, which stands for Area Concentration Achievement Test. Each year the department selects four of ten subcategories ranging from developmental or abnormal psychology to animal learning and motivation or human learning and cognition.

The selections tested take about fifty minutes. The purpose is to benchmark the different curriculum areas and see "what holes need to be filled," said Muir. "You can do it in your class. But it is nice to do it as a department."

Indeed, the results of St. Olaf's assessment plan are now used to improve teaching and learning. Departments change what needs to be improved and ask for support if needed from administrators to make the desired changes. Assessment has provided departments with feedback on student writing, for example. But faculty and administrators alike say that the process of engaging in the assessment activities has led to more formal faculty conversations. Faculty members share their best practices and collaborate both with administrators and students.

In chemistry, for example, the department identified five desired learning outcomes centered on knowledge, practice, information about the literature, written and oral communication skills, and safety. It first gathered evidence on knowledge and safety since there were existing evaluation instruments to use. But then two professors wanted to know whether they had the same outcomes. ("Luckily, they did," noted Walczak.) Both saw that there were weaknesses in their instruction of electrochemistry. Students were introduced to the topic freshman year and again in analytical chemistry junior year. Yet assessments revealed that the instructors assumed students remembered more than they did. Now the teachers delve more deeply into the subject freshman year and then do a review at the beginning of the junior-year course.

Faculty and administrators alike say that the process of engaging in the assessment activities has led to more formal faculty conversations.

Assessments also revealed other smaller, but consequential, weaknesses. After an assessment of safety practices for lab work, the faculty was surprised to learn that students didn't really know how to use different fire extinguishers. The safety manual was rewritten to make directions clearer. The point is to "start with something small," said Walczak. "There is a tendency to treat everything at once. That is overwhelming. Pick one thing and be thoughtful about how you gather evidence."

To Eric Lund, using assessment now "makes me much more conscious when we are evaluating programs or adding a program," he said. In expanding the assessment of study abroad programs, the department first used an inventory developed at James Madison University called the BEVI, an inventory of beliefs, events, and values used to analyze the study abroad experience. Today, Lund uses a range of instruments to examine the effects of programs, from measuring whether students are interested in global issues as a result of their experience to assessing what they learned about the natural world or social problems.

Students are now given an assessment before they go abroad, when they return, and then again a few months after they've returned to see what lingering effects there are from the experience.

In spite of its recent award, St. Olaf is not resting on its laurels. It was recently named one of fifteen institutions involved in a collaborative assessment project run by the Center for Inquiry in the Liberal Arts at Wabash College in Indiana. It is collaborating with other colleges through the Higher Education Data Sharing Consortium. During this year of reflection, assessment data are feeding into the college's new strategic plan. And after this year of reflection, where is St. Olaf headed in the future? Beld predicts that with "our rotating four-year schedules, including time for reflection, teaching and learning will constantly be evolving. We will build on previous activities."

University of Wisconsin–Madison:
The Wisconsin Experience Shapes Assessment

It's no surprise that the University of Wisconsin–Madison (UW–Madison) is among the recent leaders thinking about how to improve the quality of life for its students by developing new forms of assessment. The university has been a leader in progressive education dating back well over a century. The Wisconsin Idea, for example, was formulated by UW President Charles Van Hise in 1904 at the height of the "progressive era," and sought to identify the university's mission with solving social, environmental, and agricultural problems facing residents of the state.

UW–Madison, the state's flagship university, continues to draw upon the Wisconsin Idea. But as a large public research university it is now home to 42,099 students and 2,017 faculty members. It offers 5,100 courses in 20 colleges and schools with 161 undergraduate majors, 159 master's programs, 110 doctoral programs, and 12 professional degrees. The sheer size and scope of the university pose modern challenges, leading faculty and administrators to build upon its history while creating new experiences to ensure student success.

How the Twain Converged

Aaron M. Brower was one professor who built upon the Wisconsin Idea by creating new experiences for undergraduate students. As a professor in the School of Social Work, he'd focused on the policies and programs that influence undergraduates as they make the transition from high school to college. He set up residential learning communities, first-year experience programs, and then the Wisconsin Experience, a program now greeting freshmen when they arrive on campus. Along the way, he wondered about how a college integrates its campus environment with student learning and then who will be successful in adapting.

"That led me to ask, 'What do we mean by success?'" said Brower, who is now vice provost for teaching and learning, "and then how do we assess 'success.'" And that, in turn, led him to talk with people about how to create learning outcomes and assess integrative learning.

By 2007, Maureen "Mo" Noonan Bischof, who'd come out of the School of Education to become assistant vice provost, had become involved in AAC&U's LEAP initiative. Wisconsin was one of LEAP's early partner states, exploring what students throughout the system should learn as part of their liberal educations and through the clearer definition of learning outcomes. When Bischof heard about Brower's interest in creating learning outcomes, she suggested he look at AAC&U's. "The [LEAP] essential learning outcomes fit with what Aaron was thinking about," said Bischof. The two joined forces.

Meanwhile, Jocylyn Milner had been a program coordinator in the plant pathology department in the mid-1990s when "we were starting to think about learning outcomes at the program level," she said. "Working with the faculty, I wrote an assessment plan." Because she was familiar with assessment, in 1999 she moved to Academic Planning and Analysis, which includes institutional research and other academic planning activities such as program review. Since 2003 she has been its director.

Faculty and administrators alike say that the process of engaging in the assessment activities has led to more formal faculty conversations.

Milner, Brower, and Bischof combined efforts in 2007. "Our roles merged with many other people on campus," she explained. "Aaron had not been in his role as vice provost before, and he brought with him fresh ideas that evolved into something we now call 'the Wisconsin Experience.' His approach to things is very collaborative. He helped bridge us together."

The group, dubbed the Convergence Group, began to articulate learning goals for the campus as a whole. "Our focus for many years has been on academic programs," said Milner. "It became clear that there was a gap for us. We didn't have a set of articulated learning outcomes at the campus level."

So Bischof and Milner decided to audit faculty documents that talked about the student learning experience. "They didn't use the language of 'learning outcomes,'" added Milner, "but they had embedded in them these ideas. . . . It became clear that the faculty documents really spoke strongly to the essential learning outcomes and we recommended to the University Assessment Council that they be adopted as our learning outcomes. Then there was a series of conversations, governance meetings, and testing of that water. Eventually, a number of university committees affirmed them. We've had them as our framework since 2008 and we continued to build assessment into programs."

Essentially, The Wisconsin Experience and the essential learning outcomes became one document but two frameworks converging to align with the curriculum of various

schools and departments at UW–Madison. They are used to guide and evaluate teaching and learning throughout the university.

What the Wisconsin Experience Lends to Assessment

When UW–Madison students arrive on campus as freshmen, they are introduced to the Wisconsin Experience, both a concept and a program that states, "Together, we create and apply learning inside and outside the classroom to make the world a better place." That language reflects the continuity in the university's goals over the past century since the days of Van Hise.

But the contemporary Wisconsin Experience goes further to more specifically describe the "essential learning" at UW–Madison right off the bat. According to Brower, it includes "the ability to engage in the world, to be creative problem solvers, to integrate empirical analysis and passion, to seek out and create new knowledge and technologies, and to engage as world citizens."

First-year students start off in several "inquiry-based, high-impact practices including research experiences that generate knowledge and analytical skills; global and cultural competencies and engagement; leadership and activism opportunities; and application of knowledge in the real world." These are developed through a range of activities, including research apprenticeships on competitively funded projects, First-Year Interest Groups (FIGs), and national and international internships.

Students also are introduced to UW–Madison's essential learning outcomes, which grew out of the AAC&U version.

Coordinating an Assessment Plan

Bischof is now "the point person," she said, coordinating assessment and student learning under the domain of the Office of the Provost. Bischof notes it is a systematic approach and that there are three "prongs," as she puts it, to assessment:

- *A University Assessment Council.* A council for schools, colleges, and units that coordinates and supports assessment. The council includes, for example, staff members who can provide guidance on the development of surveys, analysis, and reporting and an academic technology unit so new innovations can be used.
- *The School and College Assessment Plan.* Every college has its own plan for assessment and must report annually with updates on what progress has been made and what program assessments have been done.

- *A University Assessment Fund.* Each year, individual departments can apply for seed money to devise assessments. "This helps support efforts across the campus," said Bischof. About fifteen proposals are funded each year. The funding comes from the Madison Initiative for Undergraduates, a tuition-funded effort to provide financial aid, noted Bischof, "and to enhance the undergraduate experience by funding new faculty, new programs, and curriculum development. It provides incentives to enhance undergraduate education, but attention must be paid to learning outcomes."

The history department, for example, received a grant to use assessment to reconceptualize its major and devise new assessment tools. After surveying students and recent graduates, the department found that students felt they were not well enough prepared for primary research, according to Sarah Thal, associate professor and associate chair and director of undergraduate studies. As a result, the department introduced an intermediate course to better prepare students for embarking on a senior thesis. Using funds from the Madison Initiative, the department further defined its learning outcomes and assessment tools. It restated what the history major should be, "not just the breadth and knowledge," said Thal, "but explicit skills."

Initially, Thal went to Bischof for advice. "I said, here is what we want to accomplish," said Thal. "I then came back with drafts for feedback." Bischof urged her to initially focus on measuring three learning outcomes, Thal noted, and in 2011–14, the department will be assessing research, writing, and the ability to articulate skills and their applicability.

Other Facets of UW–Madison Assessment

The essential learning outcomes also served as a critical matrix for UW–Madison in the assessment of its general education requirements. In 1994, the university had no undergraduate requirements shared by all schools and colleges, nor written or oral communications or information literacy skills shared by freshmen and sophomores. For example, less than 5 percent of UW–Madison students were required to complete a writing course. In 1996, the university required new students to complete requirements in communications, quantitative reasoning, ethnic studies, and a new "breadth" requirement.

A general education assessment council now oversees the assessment work, and the general education program now requires assessment activities. Multiple approaches are used to assess outcomes, including quantitative social science research practices (data analysis, surveys, comparison of pre- and post-tests) and qualitative approaches (focus groups, portfolio analysis).

Additionally, "every department is theoretically engaged in assessment," said Brower. "And every ten years departments must go through a review. Ten or twelve are going through a review at any one time. Some schools are reviewed by their own accrediting bodies."

UW–Madison provides online tools for faculty for a variety of assessments. One, an online assessment primer, asks the teachers of science, math, engineering, and technology, for example, to push students to "more challenging heights." It argues that traditional testing methods are too limited for guiding student learning and urges teachers first to articulate their course goals and then to use the website's Classroom Assessment Techniques that are aligned with those goals. It guides faculty in assessing student understanding, skills, and attitudes "through concept maps, concept tests, Fermi tests, interviews, portfolios," and other techniques. This website is supplemented by a team of experts who can help individual instructors drive their student learning through assessment.

The university also maintains a website on assessment for department chairs, program faculty, staff, and others involved in assessment activities. It provides guidelines, suggestions, and updates. They also have access to a UW–Madison assessment manual, forms for annual assessment reports, and assessment resources and related websites.

In addition to surveys and other instruments already administered, Milner said that she foresees working in the future with Brower and Bischof on the next step of campus-level assessment activities: collecting artifacts from student works in the disciplines and developing rubrics. These will include senior theses, art exhibits, and engineering expo projects. "Our students do such amazing work," she said, "it would be great to get a sense from them where the strengths of those programs are to improve the educational experience."

Contributors

Robert J. Sternberg is provost, senior vice president, and Regents Professor of Psychology and Education at Oklahoma State University, as well as Honorary Professor at the University of Heidelberg, Germany. Previously, he has been dean of the School of Arts and Sciences at Tufts University and IBM Professor of Psychology and Education and Professor of Management at Yale University, where he directed the Center for the Psychology of Abilities, Competencies, and Expertise. Sternberg is treasurer of the Association of American Colleges and Universities. He is past president of the American Psychological Association and the International Association for Cognitive Education and Psychology, and president-elect of the Federation of Associations of Brain and Behavioral Sciences. He is a member of the National Academy of Education and the American Academy of Arts and Sciences. Sternberg's PhD is from Stanford University, and he holds twelve honorary doctorates from eleven countries. He is the author of roughly 1,300 publications.

Jeremy D. Penn is director of university assessment and testing at Oklahoma State University. He is a member of the Association for the Assessment of Learning in Higher Education and the National Council on Measurement in Education. He is the editor of "Assessing Complex General Education Student Learning Outcomes," a volume in the *New Directions in Institutional Research* series, and has presented on assessment at regional and national conferences. Prior to joining Oklahoma State University, he worked with program assessment and accreditation at the University of Nebraska–Lincoln and taught high school mathematics. He holds a PhD in qualitative, quantitative, and psychometric methods from the University of Nebraska–Lincoln in 2008.

Christie Hawkins is director of institutional research and information management at Oklahoma State University. She is a member of the Southern Association of Institutional

Research, as well as the Oklahoma Association of Institutional Research and Planning, on whose board of directors she has served as secretary and president. Hawkins holds a PhD in research, evaluation, measurement, and statistics from Oklahoma State University.

Sally Reed is a writer and editor specializing in educational publishing, and a consultant on public policy and other issues affecting higher education. She was a senior consultant to the late Ernest L. Boyer, president of the Carnegie Foundation for the Advancement of Teaching; an associate vice president of Loyola University Chicago; and author and editor for a range of publications including Phi Delta Kappan, *The New York Times' Education Supplement*, and *Focus* for the Lumina Foundation. She was a researcher and writer on *Smart Schools, Smart Kids* (Simon & Schuster) and she's won various awards for her work from organizations such as the Education Writers Association and the Educational Press Association. She also is an editor of *COLLEGE BOUND: Issues & Trends for the College Admissions Advisor* (www.collegeboundnews.com).

About AAC&U

The Assocation of American Colleges and Universities (AAC&U) is the leading national association concerned with the quality, vitality, and public standing of undergraduate liberal education. Its members are committed to extending the advantages of a liberal education to all students, regardless of academic specialization or intended career. Founded in 1915, AAC&U now comprises more than 1,200 member institutions—including accredited public and private colleges and universities of every type and size.

AAC&U functions as a catalyst and facilitator, forging links among presidents, administrators, and faculty members who are engaged in institutional and curricular planning. Its mission is to reinforce the collective commitment to liberal education at both the national and local levels and to help individual institutions keep the quality of student learning at the core of their work as they evolve to meet new economic and social challenges.

Information about AAC&U membership, programs, and publications can be found at www.aacu.org.

Association of American Colleges and Universities

1818 R Street, NW | Washington, DC 20009
202.387.3760 | www.aacu.org

ISBN 978-0-9827850-4-1