CLOSING THE ASSESSMENT LOOP

By Trudy W. Banta and Charles Blaich

Trudy W. Banta is a professor of higher education and senior advisor to the chancellor for academic planning and evaluation at Indiana University-Purdue University Indianapolis. Recipient of eight national awards, Banta has written or edited 17 books on outcomes assessment in higher education, including, with Elizabeth A. Jones and Karen E. Black, Designing Effective Assessment: Principles and Profiles of Good Practice *(Jossey Bass, 2009). Charles Blaich is the director of the Center of Inquiry in the Liberal Arts at Wabash College. He taught at Eastern Illinois University from 1987 to 1991 and moved to Wabash in 1991, becoming the director of the center in 2002. Blaich also directs the Wabash National Study and the Teagle Assessment Scholar program.*

The idea for this article originated in what sounded like a simple request from *Change* editor, Peg Miller. She asked for some examples of the ways in which the results of student learning outcomes assessments, particularly those derived from standardized tests, had been used to stimulate improvements in teaching, learning, and student services such as advising. The request sounded reasonable—until we began searching for examples.

We scoured current literature, consulted experienced colleagues, and reviewed our own experiences, but we could identify only a handful of examples of the use of assessment findings in stimulating improvements. In fact, among 146 profiles of good practice submitted by colleagues at campuses from across the country for possible inclusion in a new book, Trudy Banta, Elizabeth Jones, and Karen Black found that only 6 percent of the profiles contained evidence that student learning had improved, no matter what measure had been used. Likewise, in their evaluation of the Wabash National Study, Charles Blaich and Kathleen Wise noted strong campus engagement with the process of assessment but few instances of actual change in response to the information generated by the study.

Accreditors, speakers at assessment conferences, and campus leaders all decry the fact that too few faculty are closing the loop—that is, studying assessment findings to see what improvements might be suggested and taking the appropriate steps to make them. This is difficult enough with locally developed measures; adding the need to interpret nationally standardized test scores and connect them with local programs and teaching approaches exacerbates the difficulty of the task. It is even rarer to find that the *effects* of making improvements on the basis of assessment findings are monitored over time to see if the desired outcomes are attained.

Many articles and books describe the qualities of good outcomes assessment. In her new book, Linda Suskie devotes

TABLE 1. CHARACTERISTICS OF EFFECTIVE OUTCOMES ASSESSMENT

Planning
- Involves stakeholders from the outset to incorporate their needs and interests and to solicit later support.
- Begins when the need is recognized and allows sufficient time for development.
- Has a plan with clear purposes that are related to goals people value.
- Bases assessment approaches on clear, explicitly stated program objectives.

Implementation
- Has knowledgeable, effective leadership.
- Involves recognition that assessment is essential to learning, and therefore is everyone's responsibility.
- Includes faculty and staff development.
- Places responsibility for assessment at the unit level.
- Uses multiple measures, thereby maximizing reliability and validity.
- Assesses processes as well as outcomes.
- Is undertaken in an environment that is receptive, supportive, and enabling.
- Incorporates continuous communication with constituents concerning activities and findings.

Improving and Sustaining
- Produces credible evidence of learning and organizational effectiveness.
- Ensures that assessment data are used continuously to improve programs and services.
- Provides a vehicle for demonstrating accountability to stakeholders within and outside the institution.
- Makes outcomes assessment ongoing, not episodic.
- Incorporates on-going evaluation and improvement of the assessment process itself.

Source: Banta, T. W. and Associates, *Building a Scholarship of Assessment.* San Francisco, CA: Jossey-Bass, 2002, pp. 262–263.

a chapter to this topic. Banta, drawing on several prior lists, identifies 17 characteristics of effective outcomes assessment (see Table 1), including beginning with a written plan with clear purposes, providing for faculty and staff development, and ensuring that assessment data are used continuously to guide improvements. Presenters at national and regional assessment conferences also provide examples of effective practices. With so much good advice available, why are improvements in student learning resulting from assessment the exception rather than the rule?

Effective Assessment

The Cognitive Level and Quality of Writing Assessment CLAQWA) was created in the late 1990s by Teresa Flateby and Elizabeth Metzger to assess student learning in a two-year learning community program at the **University of South Florida**. The CLAQWA is a sixteen-trait rubric with five levels that provides a consistent basis for faculty assessment of student writing assignments across disciplines. The cognitive portion of the scale is derived from Bloom's Taxonomy of Educational Objectives, and the writing portion applies principles from commonly used writing handbooks.

The CLAQWA rubric communicates faculty expectations clearly and provides detailed feedback to students about their strengths and weaknesses. Thus individual students know which aspects of their writing and thinking need to be improved. The CLAQWA also yields data on group strengths and weaknesses that faculty can use in rethinking their assignments and pedagogy.

Improvements over time in the scores of individual students and of student cohorts (course sections or other groupings) demonstrate that use of the CLAQWA has improved student learning. The CLAQWA is now online and has been applied by faculty in disciplines as diverse as theatre and electrical engineering.

In this article we will describe some of the conditions that make it difficult to close the loop. We discuss the importance of faculty engagement in assessment, the difficulties created by external mandates for assessment and for testing, the challenges presented by high turnover in faculty and administrative leadership, and the need to develop realistic expectations about how long it will take to move from collecting evidence to making changes. Then, in an attempt to suggest a way of addressing these concerns, we introduce the concept of double-loop learning in assessment as a mechanism for increasing the likelihood that assessment will lead to improvements in learning.

TO CLOSE THE LOOP

Engaging Faculty is Essential

Although much of the national conversation about assessment focuses on measurement issues, encouraging the use of assessment data to guide change is much more about collaborating with colleagues to decide what to improve than it is about measurement. Evidence forms the basis for these collaborations, but even the most beautifully collected and interpreted evidence will have no impact on students whatsoever unless it engages an institution's faculty, staff, governance structures, faculty development programs, and leaders.

In a recent survey of chief academic officers conducted by the National Institute on Learning Outcomes Assessment (NILOA), two-thirds of the respondents said faculty engagement in particular is a key element needed to advance assessment. While evaluating the work of individual students and informing them of their strengths and weaknesses is a process in which faculty engage routinely, taking a look at student work in the aggregate, not to mention other sources of evidence, to see

where group strengths and weaknesses are occurring and using this evidence to guide improvements constitutes a new, unfamiliar, and time-consuming activity. Targeted faculty development activities are essential in preparing faculty to conduct outcomes assessment and make appropriate use of the findings.

<div style="border: 1px solid; padding: 10px; background-color: #f4a460;">

Effective Assessment

The process of feeding assessment data back into the curriculum has become a point of pride at **Hope College**. As one example, the 2004 data from the National Survey of Student Engagement (NSSE) showed that Hope students lagged their peers in academic effort, especially regarding the number of homework hours per week.

The assessment director shared the data at an extended evening faculty meeting (with dinner provided). Faculty groups proposed possible solutions to this problem that were shared with all faculty. The assessment committee also conducted focus groups with students, and results were again shared with the community. Finally, departments were asked to dedicate one department meeting to this issue and to prepare two specific strategies that they could use to improve rigor at the department level. These strategies were shared among the faculty.

Hope's 2010 NSSE data now indicate that only 21 percent (down from 38 percent in 2003) of Hope first-years and 28 percent of seniors (down from 39 percent in 2003) indicate spending 10 hours a week or fewer (compared to 33 and 34 percent of students from peer colleges, respectively).

</div>

Most faculty are deeply committed to helping their students learn. However, because they generally perceive assessment as an externally motivated and bureaucratic process, they regard it as something that steals from the time they want to devote to students. Good assessment programs have to have "face validity" for faculty, who should be able to see how the information gathered during assessment will help them in the classroom.

While the vast majority of faculty are committed to improving the extent to which students learn in their classrooms and programs, most are also committed to being scholars. The time they spend on assessment means time away from their scholarly work. Moreover, the widely held assumption that active scholarship improves one's teaching effectiveness leads some to argue that cutting into the time they spend on their scholarship will diminish the quality of their teaching.

There is little research to support this belief, and we should not let it mask the real challenge on this point: Most assessment programs are based on the assumption that student learning is the primary responsibility of faculty and therefore time dedicated to assessment is time spent working on their primary responsibility. But if the primary responsibilities of the faculty at an institution are not clear, or the institutional value and reward systems do not align with the "student-learning-first" rhetoric, then many faculty will duck assessment responsibilities.

An important lesson from Banta, Jones, and Black and the Wabash study is that assessment evidence has to address questions that will support faculty and staff efforts to help their students learn. One effective way to build that support is to construct local measures. Another is to use campus conversations about standardized measures to identify connections between those measures and the things that, based on their work with students, faculty and staff identify as important.

How should conversations about assessment be structured? Assessment leaders should avoid doing presentations in which the data and conclusions are simply handed out to faculty. If faculty do not participate in making sense of and interpreting assessment evidence, they are much more likely to focus solely on finding fault with the conclusions than on considering ways that the evidence might be related to their teaching.

At many institutions in the Wabash Study, faculty and staff meet with Center of Inquiry staff to review and consider assessment evidence. In these meetings Center staff talk about some patterns that they see in assessment data and then ask, "What do you think this means?" The goal in these conversations is not to accept just anything that people say in interpreting the data but to engage in a "Yes, that sounds reasonable but how is that consistent with what students say on these questions?" kind of conversation. In many ways, good discussion about assessment data resembles a good seminar discussion about a book. People cite the text, in this case the data, and then dig in, push back, consider their own experience, and try to find broad themes.

One way these conversations with faculty have been dramatically improved in the Wabash study is by including information from student conversations about their experiences at the institution. When the Center of Inquiry staff and Teagle Assessment Scholars visit a campus, they typically spend the first day of their visit talking informally with three or four groups of students. These conversations are enormously informative. They allow the visitors to get a sense of the "lived" learning environment that is so much richer than any institution's documents, assessment data, or Web pages provide.

The conversations with students also enable the Center staff and Teagle Assessment Scholars to encourage students to become engaged with their institution's assessment data. Students are remarkably interested in talking about their institutions and about what is working well for their learning and what is not.

> **If faculty do not participate in making sense of and interpreting assessment evidence, they are much more likely to focus solely on finding fault with the conclusions than on considering ways that the evidence might be related to their teaching.**

The notes from these conversations then inform subsequent discussions with faculty. In addition to looking at survey data and findings on learning outcome measures, inserting themes and occasional quotes from the students into faculty conversation about assessment data allows us to move beyond the concern that "students at my institution would never understand what a survey means when it asks how many drafts they typically write before they hand in a paper" to providing details about what students may be pointing to in their survey responses.

Of course there are caveats. As Upcraft and Schuh argue, unlike research, assessment is oriented toward action, not the level of deeper clarity and precision that faculty strive for in their scholarship. The consequent need for "just one more" survey, focus group, or portfolio review may be wise, but it can also be a way of postponing action and preserving the status quo.

One of the challenges of translating assessment evidence into improvement is for assessment leaders to know when gathering more information would help focus and clarify potential actions and when their knowledge is good enough to change a class or program. The goal of assessment is not just to gather evidence, after all, but to make evidence-informed changes.

External mandates don't facilitate campus engagement.
NILOA survey respondents made it clear that regional and disciplinary accreditors provide the primary impetus for campus engagement in outcomes assessment. Accreditors have been phasing in their assessment guidelines for institutions since the late 1980s. These associations are governed by campus representatives and provide a buffer between the institutions and state and federal governments. Their influence has been gradual and generally helpful.

State mandates and reactions to national calls for accountability exemplified by the Voluntary System of Accountability (VSA) have been less gradual and thus more actively resisted, not only by faculty but by students. According to Pat Hutchings, from its earliest days, assessment became identified with actors outside academe whose patronage cast a pall over what assessment might have become within the academy.

Some of the strongest resistance has been evoked by requirements to use standardized tests of general intellectual skills. Banta, Jones, and Black found that only 8 percent of the authors of their 146 profiles of good practice mentioned such tests as one of their strategies for assessing learning. And when the profile authors described the uses made of assessment findings to guide improvement, the test scores were not mentioned. Instead, supplemental measures constructed locally and indirect evidence derived from questionnaires and interviews were identified as the stimuli for planned responses.

In work with Wabash Study institutions, Blaich and Wise encountered the same phenomenon. Scores and value-added statistics based on the Collegiate Learning Assessment (CLA) and Collegiate Assessment of Academic Proficiency (CAAP) were not productive of significant change unless they were woven into institutional conversations with students, faculty, and staff. Even when test scores were available, they did not attract as much attention as did students' responses derived from surveys and focus groups.

Nevertheless, students' performance on such tests can be revelatory, sometimes in unexpected ways. For example, one Wabash

> **The…need for "just one more" survey, focus group, or portfolio review may be wise, but it can also be a way of postponing action and preserving the status quo.**

study institution experienced a notable decline in students' scores on a standardized measure of critical thinking. It was clear both from the pattern of responses on the test items and on questions about how much effort they had expended that the decline was due to students' lack of engagement with the test.

However, it turned out that that lack of motivation in taking the test was consistent with other data on students' academic motivation, the low number of hours they reported doing academic work outside class, their high levels of alcohol consumption, and how "uncool" they felt it was to be academically engaged. Although the standardized test did not help the campus understand how its students' critical thinking was developing, the students' disengagement from the test was woven together with different threads of quantitative and qualitative assessment information to prompt an important conversation about the level of students' academic engagement.

At a number of Wabash study institutions, students' ratings declined significantly during their first year of college on two different measures of their openness to diversity: the Miville-Guzman Universality-Diversity Scale and the Openness to Diversity and Challenge Scale. This decline became an important part of faculty and staff discussions, especially after focus groups were convened to talk about the surveys with students at those institutions.

Whether institutions are building their own assessment measures or adopting standardized measures, the key is to build the connection to local campus concerns intentionally. If it is not possible to build this connection with standardized instruments, these tests and questionnaires are unlikely to be useful for promoting student learning. Even distributing a report about the results from a portfolio review or a survey will have little impact without support to help practitioners make sense of the findings and connect them with possible actions in their courses or programs. Data alone will not carry the day.

High turnover rate in faculty and administrative leadership for assessment. Banta, Jones, and Black found that 42 percent of the 146 assessment programs they studied were just two years old or less. One reason for this is that presidents and chief academic officers (provosts, vice chancellors for academic affairs, deans of instruction) generally do not stay long in these roles. When they move on to other positions, their successors are likely to have different views of assessment and thus change directions.

When the position of assessment coordinator is vacated, again the approach to assessment campus-wide may change, especially if an individual inexperienced in assessment assumes the responsibilities and must take time to become educated, then choose his or her own path. Even permanent faculty who are asked to take on the role of campus assessment coordinator often decide that they need to return to their disciplines in order to stay current or to have more contact with students.

Interestingly, changes in administrative leadership, faculty, and staff often have a more profound impact on institutional assessment work than on other functions such as admissions, assigning students to residences, scheduling courses, and granting degrees. This shows that using evidence to promote improvements is not yet a core institutional function.

Unrealistic timelines for change. Collecting and reviewing reliable evidence from multiple sources can take several years. It is too expensive to do everything—standardized test, national survey, locally developed measures—for all stated outcomes in the same year. So there must be a multi-year schedule. Yet state mandates or impatient campus leaders may exert pressure for immediate action. Pat Hutchings (2010) says that Missouri's state motto, "Show Me," captures the tone of policy makers who are tired of what they view as higher education's sense of entitlement and are asking for accountability.

But trying to force change can lead to faculty frustration and ultimately to resistance if results are disappointing—particularly if low test scores or negative satisfaction ratings are met with disapproval or even punishment by campus administrators or policy makers. Effective assessment takes time to plan, implement, and sustain. And faculty need to have reason to trust that disappointing findings will be met with offers of assistance in taking corrective action.

On the other hand, waiting for perfect data or confirmation of findings from multiple sources over multiple years must not lead to paralysis and fear of taking any action at all. Since faculty are the ones who must use assessment evidence, it is both reasonable and necessary to have them play a role in making sense of that evidence, but with the expectation that they will act on their analysis. If data seem to confirm their previous experience, faculty may be motivated to implement an improvement quickly.

Institutional Learning

Organizations, like individuals, need ongoing feedback on the impact of their efforts that comes from sources outside their day-to-day experience.

Assessment is a learning process— that is, it takes trial and error for institutions to figure out what and how to assess. Moreover, to be successful in improving student learning, assessment programs need to evolve as incoming student qualities, institutional learning goals, faculty, and resource levels change. So assessment programs themselves should be assessed continually.

A "single-loop" approach to evaluating assessment programs revises and tunes the way things are currently done: ways of sampling student work or how students are solicited to participate in a test or survey. A "double-loop" approach, as conceived originally by Chris Argyris, encourages more fundamental

questions about learning goals and whether the processes to assess them are in fact leading to improved student learning. [Editor's note: for a discussion of single- and double-loop learning, see the article by John Tagg in the July/August 2007 issue of *Change*.]

Assessing Assessment

A key step toward increasing the effectiveness of any assessment program is to engage in periodic deeper formative evaluation, in which the question is not only how well assessment tools are being deployed and whether targets for gathering and reporting on assessment evidence are being met but whether those goals reflect core values and how effective the program has been in reaching them.

The assessment of an assessment program should not focus primarily on the quality of the assessment measures but on whether the information that has been gathered by various means is the right information about the right goals, is being considered by the right people, and is leading to action.

Three key components of any assessment program should be evaluated: resource allocation, communication, and getting assessment evidence to interested users.

Resources. An effective assessment program should spend more time and money on using data than on gathering it. This means sponsoring faculty, staff, and student discussions of the data and providing support for making changes in response to the evidence. If all of an assessment program's resources are gobbled up gathering evidence, no change is likely to occur.

Communication of assessment results. The best assessment strategies cannot be effective if the data are hidden because they are too controversial or are presented in reports that are disseminated without public calls for a response from institutional leaders. If someone from the faculty, staff, or student body were randomly selected, could they identify the outcomes, measures, and recent findings of their institution's assessment program? If asked, would faculty in a department, or the director and staff of a student affairs program, name the same two or three things that their unit is doing well and cite the evidence that supports their assertions? Would they identify the same two or three weaknesses on which their unit is focusing for improvement and cite the evidence they will use to evaluate the success of their efforts to improve? If the answer to these questions is no, then the department or program needs to review and strengthen its communication plan. Units should consider their target audiences in advance of any attempt to disseminate assessment evidence, summarize the responses to a dissemination effort, or use that information to change future attempts to communicate with those audiences.

Getting evidence to potential users. It is critical that people who feel even a small interest in using assessment information be able to access it easily and to contact someone who will help them with their questions. Increasingly campuses are posting assessment evidence on internal or external sites. How often do people access these sites and how much time do they spend reading the articles? Do faculty, staff, and students know where

they can find assessment evidence to address questions about their programs, departments, or majors? If they call or email someone with questions about assessment results, do they get a response?

NILOA's scan of institutional Websites revealed that very few institutions had easily accessible assessment information posted there—that is, information that could be accessed within five clicks from the home page.

The Purpose of Assessment

An internally driven, formative approach to assessment is based on the belief that a key factor inhibiting improvements in student learning or allowing students to graduate without learning enough is that faculty and staff who deal with students lack high-quality information about the experiences and conditions that help students learn. If they had information about how much their students were or were not learning and the practices and conditions that helped them learn, practitioners would put this knowledge to work, and improvement would naturally follow.

From the perspective of those looking for accountability, assessment is beneficial because it generates knowledge for stakeholders. The assumption is that colleges and universities operate in a market, and currently too little solid information is available on which potential students can base their decisions about which college to attend. Only by holding institutions publicly accountable for evidence of learning—that is, for providing information that may encourage the willingness to select a particular institution —will we ensure that student learning is among the high-priority activities in which colleges and universities engage.

In both cases, we assume that colleges and universities are structured in such a way that information about student learning will change practice. But as Argyris points out, double-loop learning is challenging for most organizations because they unwittingly work in a way that suppresses people's abilities to deeply reflect on and improve how they operate.

Stakeholders unconsciously work to control tasks, protect their self-interests, decide how much to tell others about what is going on, minimize the risk of losing, and avoid giving bad news to people higher up in their organizations. These actions prevent the questioning of assumptions and search for incongruities in our actions that is necessary for double-loop learning and for effective action. Surely assessment programs can be revised so that the information they provide for faculty, staff, and administrators is more actionable.

But there is reason for that defensiveness: Assessment is, at its core, a subversive activity. Although most institutions operate the way they do because faculty, staff, students, and administrative leaders genuinely believe that the current structures promote learning, the current state of affairs at almost every institution is based on a delicate set of compromises and optimizations in which many parties have participated and which few care to alter. Assessment evidence can call into question long-standing agreements, priorities, and modes of practice because these do not support student learning in the ways in which the people who created them imagined. Many assessment efforts' lack of impact may be just as much about the willingness of institutional stakeholders to reconsider deeply held

assumptions about their institutions in light of evidence as it is about the quality of assessment programs themselves.

Assessment efforts must be upgraded to ensure that they are far more likely than they are at present to lead to improvements in student learning. A key step in doing so is to emphasize that the most important outcome of assessment is not gathering high-quality data, generating reports, or stimulating conversations among colleagues. That outcome is instead demonstrably improving student learning by assessing it and using the findings to revise programs accordingly. C

RESOURCES

■ Argyris, C. (1992). *On organizational learning.* Cambridge, MA: Blackwell.

■ Banta, T.W. & Associates. (2002). *Building a scholarship of assessment.* San Francisco, CA: Jossey-Bass.

■ Banta, T.W., E. A. Jones, & K. E. Black. (2009). *Designing effective assessment: Principles and profiles of good practice.* San Francisco, CA: Jossey-Bass.

■ Blaich, C. F., & Wise, K. S. (2009). *Hampshire and Wabash assessment collaboratives review.* http://www.liberalarts.wabash.edu/research-and-publications/

■ Bloom, B. (1956). *A taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive domain.* New York, NY: McKay.

■ Hutchings, P. (2010, April). *Opening doors to faculty involvement in assessment.* National Institute for Learning Outcomes Assessment Occasional Paper #4. Retrieved from www.learningoutcomeassessment.org

■ Kuh, G., & S. Ikenberry. (2009, October). *More than you think, less than we need: Learning outcomes assessment in American higher education.* National Institute for Learning Outcomes Assessment, www.learningoutcomeassessment.org

■ Miville-Guzman Universality-Diversity Scale (M-GUDS). Retrieved from http://www.liberalarts.wabash.edu/study-instruments/

■ Openness to Diversity and Challenge Scale. Retrieved from http://www.liberalarts.wabash.edu/study-instruments/

■ Suskie, L. (2009). *Assessing student learning: A common sense guide* (2nd ed.). San Francisco, CA: Jossey-Bass.

■ Upcraft, M.L. and Schuh, J.M.H. (2002). Assessment vs. research: Why we should care about the difference. *About Campus, 7,* 16–20.