

# A CULTURE OF EVIDENCE:

An Evidence-Centered Approach to  
Accountability for Student Learning Outcomes

Catherine M. Millett

David G. Payne

Carol A. Dwyer

Leslie M. Stickler

Jon J. Alexiou





Dear Colleague:

Across the country and around the globe, higher education is undergoing major changes, including dramatic shifts in student populations, new “delivery formats” such as distance learning, innovations in curriculum design, and rising financial pressure to do more with less.

These changes come amid renewed calls for greater accountability in higher education, particularly in the United States, where the national debate is focused on student learning outcomes. Institutions of higher education are understandably concerned about autonomy. But the debate creates an opportunity for the higher education community to respond with new ways to assess the learning that occurs in college; to share their findings with students, parents and other stakeholders; and to strengthen higher education’s role as fundamental to a society’s long-term success.

As a nonprofit education research and assessment organization dedicated to advancing quality and equity in education, ETS is working closely with the postsecondary community to achieve these shared goals. We are helping to examine, define, and evaluate strategies for building “cultures of evidence” through which colleges and universities can demonstrate learning outcomes. We have presented our findings in a series of “Culture of Evidence” white papers. Our new paper is titled *A Culture of Evidence: An Evidence-Centered Approach to Accountability for Student Learning Outcomes*.

In this new paper, ETS, with the help of an advisory panel of national assessment experts, presents a framework that institutions of higher education can use to improve, revise and introduce comprehensive systems for the collection and dissemination of information on student learning outcomes. For faculty and institutional leaders grappling with the many issues and nuances inherent in assessing student learning, the framework offers a practical approach that allows them to meet demands for accountability in ways that respect the diverse attributes of students, faculty and the institutions themselves.

We are encouraged by the feedback we have received from the advisory panel and by the response from the higher education community to the first two papers. We are hopeful this new framework will be of tremendous value to educators who are committed to student success and who regard the assessment of student learning outcomes as integral to their mission.

Regards,



Kurt M. Landgraf  
President and CEO  
ETS

## ACKNOWLEDGMENTS

---

When the nation turned its attention to accountability in higher education in 2006, ETS thought about how we could best tap our resources to contribute to the dialogue. The initial decision was to write one white paper on the general topic of student learning outcomes with the higher education and assessment communities as our target audiences. As we begin 2008, we are completing what turned out to be a three-paper series. In many respects, the evolution of our series paralleled the national experience.

In 2006, we took a macro look at the current conditions. That first report, *A Culture of Evidence: Postsecondary Assessment and Learning Outcomes*, provided an overview of the assessment landscape and made a number of recommendations for addressing the dearth of empirical data on student learning in higher education.

In 2006 and 2007 we presented a second report, *A Culture of Evidence: Critical Features of Assessments for Postsecondary Student Learning*, in which we took a detailed look at current, state-of-the-art commercially available assessments. That report reviewed the major tools in use today for assessing student learning and student engagement — important aspects of the educational environment. Our intent was to provide policymakers, national organizations, and two- and four-year college and university presidents and provosts with a 30,000-foot overview.

In 2007, we turned our attention to thinking about how we could contribute to the efforts about to take place on many campuses to capture and present what students are learning. We particularly wanted to think about how campuses could capitalize on the assessments they had already been conducting while soundly preparing and positioning themselves for a world in which the demand for comparative data is increasing. This third report, *A Culture of Evidence: An Evidence-Centered Approach to Accountability for Student Learning Outcomes*, addresses both the need to measure the unique aspects of learning that occur in individual institutions and also the types of common learning that are expected to occur across all higher education institutions.

After the first white paper, we convened a 13-member advisory panel. Their contribution to our work for the second and third papers was invaluable. As a collective group, the panel was a thoughtful and critical sounding board for our ideas. We appreciate the time, insight and energy that panel members contributed to this effort. The panel helped to keep us on track to provide a series of papers we hope will be useful to a wide and diverse audience in higher education.

## ADVISORY PANEL MEMBERS AND AFFILIATIONS

---

| NAME                     | AFFILIATION  |
|--------------------------|--|
| Tom Bailey               | Professor, Teachers College at Columbia University   |
| Trudy Banta              | Vice Chancellor for Planning and Institutional Improvement, Indiana University-Purdue University, Indianapolis     |
| Douglas F. Becker        | Vice President of Development, Education Division, ACT, Inc.   |
| Steven D. Crow           | Executive Director, The Higher Learning Commission   |
| Peter Ewell              | Vice President, National Center for Higher Education Management Systems (NCHEMS)                                   |
| Steve Klein              | Director of Research and Development, Council for Aid to Education (CAE)   |
| Paul Lingenfelter        | President, State Higher Education Executive Officers (SHEEO)   |
| George Mehaffy           | Vice President for Academic Leadership and Change, American Association of State Colleges and Universities (AASCU) |
| Margaret Miller          | Director of the Center for the Study of Higher Education, Curry School of Education at the University of Virginia  |
| Liora Pedhazur Schmelkin | Senior Vice Provost for Academic Affairs and Dean of Graduate Studies, Hofstra University                          |
| Pamela E. Scott-Johnson  | Professor and Chair of Psychology, Morgan State University   |
| David Shulenburg         | Vice President for Academic Affairs, National Association of State Universities and Land-Grant Colleges (NASULGC)  |
| Belle S. Wheelan         | President, Commission on Colleges of the Southern Association of Colleges and Schools                              |

During the course of writing the third report, we benefited from the assistance of many ETS colleagues. We are grateful for the invaluable feedback of Will Jarred, Cathrael Kazin and Linda Tyler. Evelyn Fisch played a major role in coordinating the editing of the various paper drafts and scheduling our meetings with the Advisory Panel. We also want to thank our ETS colleagues in the Communications and Public Affairs Division for working with us to publish and disseminate the series.

## TABLE OF CONTENTS

---

|  |    |
|--|----|
| Introduction .....   | 1  |
| U.S. Accountability: A Summary of the Past Decade —<br>A Strong Foundation on Which to Build.....            | 2  |
| Overview of the Culture of Evidence I and II Reports.....  | 4  |
| Evidence-Centered Design as a Basis for Assessment in Higher Education:<br>Concepts for Decision Makers..... | 5  |
| Seven Steps in Creating an Evidence-Based Accountability<br>System for Student Learning Outcomes.....        | 10 |
| Conclusion.....  | 18 |
| Annotated Bibliography.....  | 20 |
| References .....   | 23 |

## LIST OF FIGURES

---

|  |    |
|--|----|
| Figure 1: Claims, Evidence, and Assessment in Evidence-Centered Design. .... | 5  |
| Figure 2: Evidence of Student Learning Outcomes Model.....                   | 10 |

## INTRODUCTION

---

The simple answer is there is no commonly used metric to determine effectiveness — *defined in terms of student learning* — of higher education in the United States.

(Dwyer, Millett, & Payne, 2006)

It seems to many observers that the walls that created silos in U.S. higher education may be starting to come down. The country may well be on the verge of creating a new paradigm for discussing and acting upon some of the most critical issues facing U.S. higher education. These issues, which have been part of the dialogue in the higher-education community in Washington, D.C., and on campuses across the country, are well-documented:

- The Organization for Economic Co-operation and Development (OECD) *Education at a Glance 2007* report shined a spotlight on the U.S. drop in rankings for higher-education attainment of 25- to 34-year-olds. The U.S. barely made the top 10 list with a 10th-place ranking.
- The National Academies' report, *Rising Above the Gathering Storm: Energizing and Employing America for a Brighter Economic Future* (Committee on Science, Engineering, and Public Policy, 2007), stressed that other countries are gnawing away at the U.S.'s long-standing pre-eminence in the global marketplace and in the science and technology arena. The committee advocated development and implementation of a comprehensive and coordinated federal effort to bolster U.S. competitiveness and restore our pre-eminence in these areas.
- ETS's report, *America's Perfect Storm: Three Forces Changing Our Nation's Future* (Kirsch, Braun, Yamamoto, & Sum, 2007), outlined the manner in which divergent skill distribution in the U.S. population, the changing global economy, and demographic trends are converging to create a number of important educational and policy challenges for the U.S.
- The Business Roundtable's (2005) work with 15 other businesses that produced *Tapping America's Potential: The Education for Innovation Initiative* challenged the U.S. to double the number of science, technology, engineering, and mathematics graduates with bachelor's degrees by 2015.
- The American Institutes for Research report, *The Literacy of America's College Students* (Baer, Cook, & Baldi, 2006), reported that more than 75 percent of students at two-year colleges and more than 50 percent of students at four-year colleges do not score at the proficient level of literacy. These students lack the skills to perform mid-level literacy tasks, such as identifying a location on a map or calculating the cost of an office supplies order.

With these different issues coming to the fore, it is not too far-fetched for people outside the higher-education community to ask, *What are students learning on college campuses today?* The response of higher-education leaders has been that there is no quick answer to the question. In the short term, institutions and researchers have provided responses based on the evidence that can be quickly generated — graduation rates, pass rates on licensure exams, and performance on graduate and professional school admissions exams. But these convenient data points do not fully answer the question of what students are learning.

In the quest to answer this question, the ongoing dialogue and debate over higher-education effectiveness has actually helped to underscore one troubling fact: the United States does not have

one metric, or even a handful of common metrics, that could paint a picture of the accomplishments of its more than 2,500 four-year and 1,600 two-year postsecondary institutions (National Center for Educational Statistics, 2006). The only relevant information is the data individual colleges and universities collect themselves and then elect to share with various stakeholders (e.g., students, parents, legislators, and accrediting agencies).

This report is intended to address the need to measure the unique aspects of learning that occur in individual institutions, as well as the types of common learning that are expected across all higher-education institutions. In the first section, we review the contributions that many organizations have made that inform our work. In the second section, we provide an overview of two earlier issue papers published by ETS. In the third section, we discuss an evidence-centered design approach to assessment that provides a useful conceptual framework for viewing the challenges within the accountability paradigm. This section discusses important issues such as validity and reliability and does so in a non-technical manner. The fourth section describes a seven-step process that can be used to guide institutions that are either re-thinking their approach to assessing student learning outcomes or are just beginning the process of considering how to develop these systems.

This report is intended to be of use to those faculty and administrative leaders who are non-experts in the assessment field, but who will be responsible for leading their institutions and/or systems through the shift toward a transparent system of accountability for student learning outcomes. We take as a given that institutions believe there is value in assessing student learning and using the results to advance important institutional and student goals. We hope to provide some insights into the conceptual and organizational challenges institutions will face as they grapple with this paradigm shift.

## U.S. ACCOUNTABILITY: A SUMMARY OF THE PAST DECADE — A STRONG FOUNDATION ON WHICH TO BUILD

---

For more than a decade, various groups have devoted their time and resources to thinking about accountability in higher education. Some of the notable work that has been carried out includes the National Association of Independent Colleges and Universities' (NAICU) work in the 1990s that produced the 1994 report *The Responsibility of Independence: Appropriate Accountability Through Self-Regulation*; the Association of American Colleges and Universities (AAC&U) Greater Expectations initiative from 2000 to 2006, which produced the 2002 report *Greater Expectations: A New Vision for Learning as the Nation Goes to College*; the Business – Higher Education Forum's working group that produced the 2004 report *Public Accountability for Student Learning in Higher Education: Issues and Options*; and the State Higher Education Executive Officers (SHEEO) National Commission on Accountability in Higher Education, which produced the 2005 report *Accountability for Better Results: A National Imperative for Higher Education*.

More recently, the Commission on the Future of Higher Education, created by U.S. Secretary of Education Margaret Spellings, contributed to the student learning outcomes dialogue. (See "A Year Later, Spellings Report Still Makes Ripples," *The Chronicle of Higher Education*, Sept. 28, 2007, for a brief review of the Spellings Commission and its impact on accountability.) The Commission's 2006 report, *A Test of Leadership: Charting the Future of U.S. Higher Education* (U.S. Department of Education, 2006), singled out student learning as one of the six pressing challenges confronting U.S. postsecondary education. The Commission urged the higher education community to incorporate student learning outcome measures that would be comparable across institutions.

The U.S. Department of Education has taken the next step. It has begun to fund efforts to move the talk into action. A recent grant to AAC&U will support a new initiative, *Rising to the Challenge: Meaningful Assessment of Student Learning*. AAC&U, along with the American Association of State Colleges and Universities (AASCU) and the National Association of State Universities and Land-Grant Colleges (NASULGC), will form a consortium to collectively build campus leadership and the capacity to implement meaningful student learning assessment approaches. They will use assessment results to improve levels of student achievement.

One of the more recent initiatives in the student learning outcomes area involves the two largest organizations of public colleges and universities, NASULGC and AASCU. These organizations and their member institutions have developed a Voluntary System of Accountability (VSA). One of its goals is to measure a set of core learning outcomes that include critical thinking, analytic reasoning, and written communication. Other organizations within higher education have also expressed an interest in assessing student learning outcomes. For example, in September 2007, a group of colleges that operate online and focus primarily on adult education announced a Transparency by Design initiative that aims to assess students in both program-level and general education-level areas (Presidents' Forum of Excelsior College, 2007).

The notions of transparency and accountability for student learning outcomes represent a paradigm shift for higher education. Although higher education has been evaluating student learning for centuries, the current societal pressure for data beyond course grades, degree granting rates, and similar "production" measures represents a sea change of considerable magnitude. (For an excellent overview of the history and uses of assessment in higher education, see Margaret Miller, "*The Legitimacy of Assessment*," *The Chronicle of Higher Education*, Sept. 22, 2006.)

Institutional leaders who understand the importance of having their colleges and universities at the forefront of this movement will need to be effective change agents within their institutions. As John P. Kotter argued in *Leading Change* (1996), one requirement for being an effective organizational change agent is the ability to create a vision and strategy for the organization. Creating a vision of what an institution will look like once it embraces the idea of transparency and accountability for student learning outcomes requires a sense of the essential elements of an effective system of assessments. This report is intended to assist institutional leaders who are responsible for leading the institutional changes that will result in greater transparency and accountability for student learning outcomes.

A range of issues arise in the context of assessing student learning and a great deal has been written about them. For example, some (Education Commission of the States, 1998; Frye, 1999; Labi, 2007) have debated whether assessment of student learning for accountability purposes is consistent with — or at odds with — assessment to improve student learning. Others have discussed whether standardized tests can truly be used to assess the learning that takes place within individual institutions, which have their own unique mission, student population, and resources. (Bollag, 2006; Eubanks, 2006; Garcia & Pacheco, 1992; and Schagen & Hutchison, 2007). Still others have discussed whether student learning should be measured in terms of general competencies (e.g., critical thinking, written communication) or in terms of skills, knowledge, and abilities within a student's academic discipline.

These issues are important. They deserve, and will undoubtedly receive, further attention within the Academy. We have included a short annotated bibliography at the end of the report that summarizes some of the key research. This report, however, is not designed to shed light on these issues. Rather, it is intended to help educational leaders, including administrative and faculty leaders, trustees, legislators, and others who are interested in asking important questions about what is required for assessing and improving student learning in a way that is transparent and that allows for the types of accountability that are being demanded in the United States.

This report is aimed at providing insights into two general issues facing institutions that wish to develop and/or refine processes and systems for assessing and improving student learning. The first issue is the factors that must be considered when asking questions about what assessments can and cannot tell us about student learning. The second issue concerns taking the steps that ensure an institution has fully engaged in the analytical work necessary to guarantee that the institutional investment in assessing student learning will bear meaningful fruit.

It is important to acknowledge some aspects of the historical context for this new paradigm of accountability for student learning outcomes. Measuring student learning outcomes is certainly not new for community colleges and four-year colleges and universities. For several decades, as part of their voluntary participation in national or regional periodic accreditation reviews, higher education institutions have worked very creatively and energetically to develop student learning outcomes measures at the campus level. These student learning outcomes measures are, in most cases, designed to reflect each institution's unique mission, curriculum, and student needs. They have tremendous potential for use within an institution. What differs between the more recent calls for accountability and local approaches and locally developed measures are the dimensions of standardization and comparability.

At its core, the issue is the extent to which inferences can be drawn about institutional effectiveness and what comparisons can be made between institutions. Locally developed measures can be extraordinarily useful in guiding important improvements in instruction and creating more effective learning environments. But unless the same measures are used across multiple institutions, it is nearly impossible to draw any meaningful conclusions about how different institutions compare in the amount or types of learning taking place.

## OVERVIEW OF THE CULTURE OF EVIDENCE I AND II REPORTS

---

This report is the third in a series of issue papers from ETS on the critical subject of assessing student learning outcomes. As a nonprofit organization with a long-established social mission to advance quality and equity in education, ETS is committed to furthering the national dialogue on how institutions of higher education (community colleges, colleges, and universities) can demonstrate accountability for their students' learning.

In the first report in this series, *A Culture of Evidence: Postsecondary Assessment and Learning Outcomes* (COE I), we described the national landscape in postsecondary assessment and learning outcomes (Dwyer, Millett, & Payne, 2006). Although there have been a number of important developments since that first report, it is still fair to say there is not a great deal of national data currently available from most institutions of higher education based on assessments of their students' learning (Curriss & Lingenfelter, 2005).

In the second report in this series, *A Culture of Evidence: Critical Features of Assessments for Postsecondary Student Learning* (COE II), we reviewed 12 major tools in use today for assessing student learning and engagement (Millett, Stickler, Dwyer, & Payne, 2007). In this high-level overview, we applied a coherent and consistent framework and language to create a guide to the most prevalent assessments of student learning. For each assessment, we provided information on the intended population, items and forms, level of results, scores yielded, comparative data availability, cost, testing sample, pre- and post-testing, time required, and annual volume or institutional data pool. We concluded the report by encouraging colleges and universities to begin their discussions about assessing student learning by asking questions, rather than simply selecting assessment tools.

It is clear from this second report that in measuring postsecondary student learning outcomes there are choices institutions can make to tailor their assessment programs to their unique missions, academic programs, and students. Each of the assessments listed in the second report offers all of the benefits of standardized tests (e.g., reliable, normative data) and they may be useful for institutions interested in assessing student learning outcomes. It is also likely that institutions will wish to employ measures in addition to, or in place of, standardized measures. This issue is addressed in the next section.

## EVIDENCE-CENTERED DESIGN AS A BASIS FOR ASSESSMENT IN HIGHER EDUCATION: CONCEPTS FOR DECISION MAKERS

The most fundamental consideration in assuring the quality of any assessment is validity. Validity refers to the degree to which evidence supports the interpretation of test scores. Evidence-centered design (ECD) is an assessment framework intended to ensure validity by aligning the assessment products and processes with the goals of assessment (Mislevy, Almond, & Lukas, 2003; Snider-Lotz, 2002). Put another way, assessment program designers can use evidence-centered design to link their decisions about the students being assessed to the information institutions need to have to support those decisions. For this reason, articulating the desired student learning outcomes is the first and most important step in the assessment process. The following discussion focuses on a conceptual understanding of evidence-centered design and how to use it, rather than on statistical issues.

Once the important student learning outcomes have been determined and the evidence necessary to collect is identified, institutions can design or select assessment activities and tools based on how well they provide the evidence. Figure 1 represents this relationship graphically.

Figure 1: Claims, Evidence, and Assessment in Evidence-Centered Design



### Claims — what is important for an institution and its students?

As Figure 1 illustrates, the first step in designing an assessment program using the principles of evidence-centered design is to specify the claims you are interested in making about the individuals

being assessed. In every case, the claims will be specific to an institution, and should be determined by those most familiar with the institution's mission and values. This means potential claims must be carefully examined by those who will oversee the assessment system and those who will be affected by them. This is a critical point, whether an institution is developing its own assessment system, reviewing one it already has in place, or selecting an assessment tool developed by others. Any assessment tool is only valid and useful when it is carefully aligned with the claims the institution is certain it wants to make. These claims may be the same as or differ substantially from those of other institutions, or from those implicit in commercially available tests.

In the case of postsecondary student learning outcomes, these claims might reflect a college's or university's formal institutional mission. They might also describe the knowledge and skills that the college or university aims to promote in students, whether or not these have been formally stated.

### **What evidence is needed to support your claims?**

Once assessment designers have identified the claims to be made about student learning, the next question that must be addressed within the evidence-centered design framework is what specific pieces of evidence are required to support each of these claims. Each claim will rest on carefully and purposefully chosen pieces or types of evidence, as well as logical explanations of how and why the evidence fits the institution's claims.

For example, an institution may need evidence that exiting students have mastered the professional knowledge in their fields of study, can perform the tasks and procedures necessary for their desired employment, and have successfully attained positions or advanced in their professional fields. The first two pieces of evidence will likely include several domains of professional knowledge and skill, respectively. The third reflects and validates the standard of professional learning the institution has claimed. That is, if a student has in fact developed the knowledge and skills in, for example, information technology (IT) necessary to attain an IT position or to advance an IT career, then that student should be able to attain such a position. Identifying a standard of performance or improvement at this stage is an important issue. It clarifies the meaning and usefulness of assessment results, so that the evidence collected is aligned with institutional goals and is appropriately interpreted by decision makers and stakeholders as supporting institutional claims about student outcomes.

### **What types of assessment tasks are best?**

Again, there is no one-size-fits-all answer. Every type of assessment has its own strengths and limitations, and particular situations for which it is more or less effective. Once assessment designers have articulated the constituent domains of each piece or type of evidence required (which will vary, at least in part, by area of study) and have justified standards of performance, they will need to consider what level of assessment complexity is appropriate. Some assessment goals require students to perform complex tasks in order to demonstrate their ability to apply their knowledge and skills to real-world problems. The performance or product of these assessments demonstrates that students can do something and are flexible to the extent that there may be various acceptable ways to approach or complete the performance or product (Messick, 1994).

Assessments can vary considerably by the amount of structure provided and the level of intricacy of the assessment product or performance. Within an evidence-centered design framework, these decisions are guided, once again, by the claims and evidence necessary to make decisions about students. Colleges and universities that need evidence about students' skills in their areas of study or chosen professional fields may base decisions about the levels of structure and complexity of assessment tasks on factors such as the nature of the subject area, the specificity of the claims they want to make about student learning, and the desired use or uses of assessment results.

More structured tasks are, in general, more appropriate for skills for which there is only one or a small number of clearly correct responses, or for claims requiring more fine-grained evidence of student learning. Less structured tasks tend to be more suitable for measuring skills for which there are many acceptable ways to complete the task or solve the problem. For example, a student studying information technology might complete a structured task to demonstrate skill in programming in a particular language. Evidence of that student's web design skills, however, might be better assessed using a less structured task that provides the flexibility to use one or many equally viable approaches to programming, organizing, and designing a website.

The web design example also represents a more complex task, requiring integration of many different skills to produce a satisfactory response. There is no one "right answer." The results of a complex assessment task such as this can be used to make claims about whether students have mastered a particular set of skills. Because so many different skills must contribute to the success of a complex task, however, the results cannot be used directly to create detailed plans for educational interventions or instructional improvements because there are just too many component skills involved to determine which ones caused a particular student's difficulty and what that student needs to do to succeed. The extent to which institutional assessment users need to know what students can do, and their recognition that competence can be demonstrated in multiple ways, will determine whether more labor-intensive and costly performance-based assessments may be warranted. Assessment complexity will likely need to be weighed against feasibility considerations, with an eye toward getting the most out of the assessment tools and activities selected.

*Selecting assessment tools.* The final step of an evidence-centered design approach to the design (or, occasionally, the selection) of an assessment program is to identify the specific assessment tools or activities that will provide the evidence needed. After carefully characterizing the evidence necessary to support claims about student outcomes, selecting or designing assessment tools and activities follows almost naturally. Because assessment program designers who have used evidence-centered design will at this point know exactly what the resulting performance, product, or item responses (that is, the evidence) should look like, setting activity parameters or choosing established tests with appropriate items is a straightforward process.

Continuing with our earlier example, colleges and universities that need evidence of an IT student's mastery of professional knowledge might review several commercially available measures of this domain and select one that is most closely aligned with the professional standards in the field, thus best reflecting the knowledge that the student needs to land a "dream job." The academic department might design, implement, and evaluate the efficacy of a variety of more and less structured tasks to provide evidence of students' skills in providing technical support, programming, and website design. Finally, assessment program designers might structure or refine processes to follow up with alumni to determine whether they were able to attain positions or advance their careers in their chosen fields, and perhaps how long it took them to do so and how satisfied they are with the roles and responsibilities of the positions they attained.

*Building in assessment validity.* Each step of the evidence-centered design process narrows the field of potential assessment tools and activities to help assessment program designers focus on those that are the most appropriate and well-suited for their purposes, while guarding against two major threats to validity (Messick, 1994): construct-irrelevant variance (measuring things you do not want to measure) and construct under-representation (not measuring parts of what you would like to measure).

This approach is highly consistent with modern views of construct validity, and with standards for ensuring the technical quality of assessments (American Educational Research Association, American Psychological Association, & National Council for Measurement in Education, 1999). By determining the specific evidence necessary to make accurate decisions about the students being

assessed, assessment designers identify and are able to implement processes to control for the influence of irrelevant aspects of assessment activities and tools on examinees' performances.

Aspects of assessments that are not directly related to the specific knowledge or skills being assessed, but lead to differences in examinee performance, are sources of construct-irrelevant variance (Messick, 1994) that necessarily weaken the inferences that can validly be made from assessment information, and can also make the assessment unfair to groups and individuals.

For example, any assessment that requires examinees to construct essays or other written responses will necessarily measure writing skill, regardless of whether making a decision about an examinee's writing proficiency is the purpose of the assessment. Even if those charged with assessing higher-education outcomes are also interested in information about writing skill, gathering it simultaneously with subject-area knowledge and skill in the student's major field can lead to ambiguous evidence. Did the student in our earlier example do poorly because he or she has not mastered the IT curriculum, or because of poor writing skills? Perhaps even more importantly, what decisions can educators and administrators make about educational interventions and program improvements to support current and future students when assessment activities and tools permit other, unintended factors to enter into the assessments? The evidence yielded may not clearly identify the knowledge and skills students have mastered and those they have not. This makes it difficult for institutions to specify clearly what their contributions to student learning outcomes are, and to know which program areas need to be improved.

The corresponding threat to validity is construct under-representation. The issue is whether the assessments provide adequate coverage of the content and processes the users are making inferences about (Messick, 1994). Notice that we use the plural, *assessments*, to emphasize that complete construct coverage need not, and in fact should not, derive from a single assessment activity or tool. The reliability and specificity of assessments is improved by increasing the number of tasks, and by triangulating multiple measures of student learning outcomes. In the examples given above, multiple assessments and other existing sources of evidence are used to establish support for the claims institutions seek to make about their students' learning outcomes.

*Validity and reliability.* The evidence-centered design framework described above makes it clear that the validity of inferences made on the basis of an assessment can no longer be conceived of as simply a correlation coefficient. Instead, it is a judgment, over time, about the inferences that can be drawn from assessment data. Reliability studies provide important data about the validity of assessments, answering questions about the consistency and stability of scores. Some examples of challenges to valid score interpretation that can be addressed by reliability studies include:

- assessments that yield dissimilar scores when the same person is retested on the same test at a different time
- different forms or editions of an assessment that have low correlations with one another
- questions or tasks within the same assessment that are intended to measure the same thing, but have low correlations with one another
- scorers or raters who have low levels of agreement with one another about an individual's performance on a question, essay, or other task

It is important to remember that unreliability is only one type of threat to validity, and it is not accurate to think of it as being more "objective" than other evidence about validity.

It is also important to remember that some extremely simple assessment formats may be highly reliable, but at the same time may be poor at measuring the knowledge and skills an institution considers critical. In such cases, it would be a poor trade-off to sacrifice a basic aspect of validity — how much of what you want to measure is actually being measured — in order to obtain information that is highly replicable but does not provide good evidence for the claims that you want to make about student learning outcomes.

As noted in COE II, test publishers have an obligation to be very clear about what reliability information they have, how it was obtained, and how it relates to the design of the assessment. The assessments reviewed in COE II represented a wide variety of approaches to reporting reliability, including measures of a test's internal consistency, measures based on a test/retest design, correlations between tasks, and correlations between scores given by different readers. Most of the data we reviewed appeared strong; and with a few exceptions, the values might seem impressive. But the more important consideration for potential users is whether the data provided are really addressing their specific concerns about an assessment's suitability for their student population and assessment needs, and whether the data that are generated from an assessment are free from contamination by irrelevant factors (such as reader or specific-task variation) that would impair the user's ability to make valid inferences from it.

As a general rule, a test with important implications for an individual is expected to be more reliable than a test with less important implications. Reported reliability estimates of .85 to .90 or higher are common in high-stakes, standardized testing. In addition, all other things being equal, longer tests are more reliable than shorter tests. So, care should be taken in interpreting tests that have been abbreviated to make them quicker to administer and subscores that are based on a very small number of questions or tasks. Reliability data in the range of .60 are not uncommon in such subscores. Essays and other constructed-response tasks often have lower indices of reliability than do multiple-choice tests. It is appropriate with such assessments to provide information on how closely the raters or scorers agree about what the score should be.

When considering using an assessment, one should look for a complete discussion from its publisher of what was done, what the results were, and why this was a logical approach given the nature of that assessment. Providing a simple correlation coefficient, no matter how high, is not an adequate demonstration of reliability and does not provide good evidence to support the claims that the institution has used as the foundation of its assessment activities.

*Practical questions for institutions to consider.* Much has been written about assessment and accountability in postsecondary education. The above discussion of using the evidence-centered design process to establish validity and reliability should be a useful tool for institutions and institutional leaders. In addition, however, we would like to suggest a number of specific questions that institutions frequently face when making or choosing assessments, place them in the context of the evidence-centered design framework, and make some observations about them.

*Are all students tested, or just a sample?* Are the samples large enough to support the generalizations that are desired? Sampling rather than assessing all students is often practical as a cost-saving or time-saving measure. Sampling means, however, that there will not be individual student scores, and that claims cannot be made about individuals. When samples are used, care must be taken to avoid compromising validity by assessing a sample of students that is, for example, unrepresentative of the other students you are making claims about, or too small to give a true idea of the institution's actual success in producing student learning outcomes.

*Are students motivated to perform in the assessment?* A basic assumption of almost every educational assessment is that test takers are providing an accurate picture of their knowledge and skills. This implies that test takers are motivated to address the tasks to the best of their ability. When this assumption is incorrect, inferences made about the assessment data are correspondingly invalid. That makes motivating students to produce their best efforts an important consideration. It is made more difficult when there are no consequences of the assessment for the individual student, or no individual results given.

*Are the assessments reviewed on a regular basis?* Circumstances change. The fundamental bases for evidence-centered design are the institution's decisions about what is important to collect data about, what the evidence should look like, and how to collect it. Any of these elements of the process can change over time, with experience, new student demographics, and shifting

institutional priorities. For these and a host of other reasons, any assessment system should be reviewed regularly.

*Is there comparability across years in the assessments used?* In addition to considering whether assessment systems are still achieving their basic goals, institutions should also consider the impact of changes made by assessment publishers on the assessments they are using.

*Are benchmarking data available?* Very often, an institution of higher education will have a number of “peer” institutions to which it would like to compare itself. When this is the case, a key consideration in the decision to develop or purchase an assessment is what assessments these peer institutions are using and whether that assessment data can be made available to you. As a general rule, it is difficult to make accurate comparisons across different tests. Again, the guiding principle should be the ability of an institution to have its claims adequately supported by the available evidence.

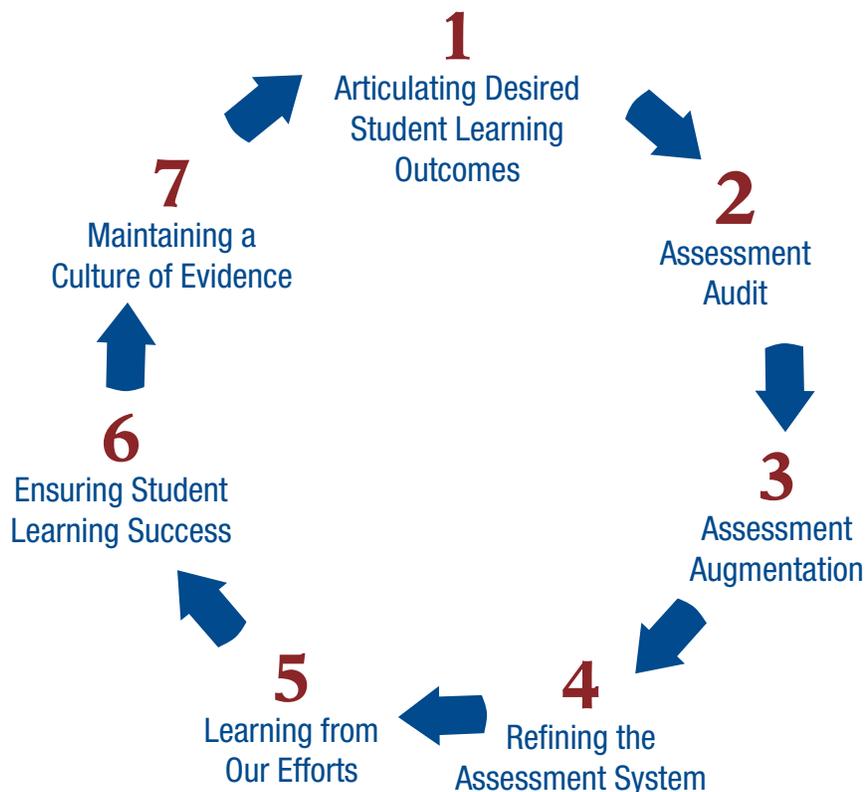
**Summary.** The evidence-centered design approach provides a rich context and conceptual framework for considering assessments of student learning outcomes and for asking important questions about the types of claims that can be made based on assessments.

## SEVEN STEPS IN CREATING AN EVIDENCE-BASED ACCOUNTABILITY SYSTEM FOR STUDENT LEARNING OUTCOMES

---

In this section of the paper, we will introduce a seven-step model that allows institutions to engage in a systematic review and analysis to help create or refine a culture of evidence for assessing student learning outcomes. Figure 2 represents the seven steps.

Figure 2: Evidence of Student Learning Outcomes Model



## 1. Articulating Desired Student Learning Outcomes: What Are Our Aspirations for Our Students to Achieve and for What Purposes Do We Wish to Document the Results?

The institution needs to determine *who should participate* in the discussion. Faculty are likely to be the group that heads the list. Other participants could include academic deans and administrators, graduate student teaching fellows, academic advisers, institutional researchers, and students.

A critical first step in creating a student learning outcomes system is to address two key issues. From an evidence-centered design perspective, Step 1, articulating desired learning outcomes, is the foundation on which all other efforts will be built. For that reason, it is important that the institution engage fully in this step.

Given the foundational nature of Step 1, it is important that we examine this step in some detail.

First, the institution needs to *define precisely what claims the institution wishes to make* about student learning outcomes. This first issue can be framed as follows:

When we take into consideration our mission, our student population and their goals and needs, what do we hope the students will achieve while they are enrolled in our institution?

These aspirations can vary widely depending upon a host of factors, including the type of institution (e.g., two-year vs. four-year; public vs. private; open admission vs. highly selective) and its location (urban, suburban, or rural). There are clearly no “correct” answers to these aspirational questions. What is important is that the institution’s faculty, administrators, and other stakeholders agree on the important questions that need to be answered. We will return to the issue of claims about student learning after we consider the second important facet of Step 1.

The second important issue that needs to be addressed involves the planned uses for the student learning outcomes data (e.g., whether the data will be used to report institution-level accountability metrics or to guide improvement of instruction within targeted courses). Although data may be used appropriately for more than one purpose (e.g., data that are intended to guide curricular improvements may also be “rolled up” to summarize institutional progress), there are limitations to the appropriate inferences that may be drawn from data. These limitations will become apparent when the institution itself determines how it intends to use the student learning outcomes data.

Taken together, the answers to these two sets of questions will help the institution evaluate the adequacy of current and planned data-collection activities and the resultant data. These two sets of issues can also help surface concerns among the institution’s constituencies. For example, faculty will undoubtedly be interested in knowing how the administration plans to use student learning outcomes data. Faculty will most likely want to distinguish assessments developed or used as diagnostic tools to identify teaching and learning opportunities from assessments that are more summative in nature and used to make peer-institution comparisons.

Returning to the claims the institution may wish to make about its students’ achievements, a number of approaches and frameworks can be used to address this issue. First, the goals and claims about postsecondary institutions’ contributions to student achievement might reflect a community college, college, or university’s formal mission. Second, the claims might describe the knowledge and skills that the institution aims to promote in students, whether or not these have been formally stated. For example, a community college serving students with diverse backgrounds and academic interests might informally align student needs with desired outcomes to generate claims such as proficiency in English-language skills, transfer rates, and success in transitions to the workforce.

These outcomes could be stated as:

1. Graduates/completers have developed the necessary academic knowledge and skills — generally and in their major area of study — to successfully transfer to a four-year institution.
2. Graduates/completers have developed the knowledge and skills in their fields of study required for successful employment and/or advancement and promotion.
3. Non-matriculated students in personal development courses have been satisfied with the content and quality of the courses they have taken.

The third approach to the issue of articulating desired student learning outcomes involves the aspects of student learning on which the institution wishes to generate evidence. The four dimensions of student learning that were articulated in COE I provide a good starting point for this approach. The four dimensions are:

- Workplace readiness and general education skills, such as the abilities to communicate clearly and effectively and to break down and analyze complex information to solve problems
- Domain-specific knowledge and skills in students' major areas of study
- "Soft skills," such as teamwork, communication, and creativity
- Student engagement with learning

Once assessment designers have identified the claims to be made about student learning, the next question that must be addressed within the evidence-centered design framework is what specific pieces of evidence are required to support each of these claims. Each claim will rest upon carefully and purposefully chosen pieces or types of evidence.

At this point, those involved in institutional assessment programs will want to carefully examine the evidence that is currently available as the result of ongoing assessment practices to determine whether this information can provide appropriate and sufficient evidence of the institution's articulated student learning outcomes.

## **2. Assessment Audit: What Existing Evidence Can Address These Student Learning Goals?**

It is safe to say that every accredited higher-education institution in the U.S. today has collected a significant amount of data of various types regarding student learning that is potentially available for analysis. For example, student grades, scores and passing rates on licensure and certification examinations, surveys of students in capstone courses, surveys of student engagement, and other types of evidence are collected at campuses across the country every year.

Regardless of whether an institution is just beginning to think about assessing student learning outcomes or is well on the way to having a system that could serve as a model for comparable institutions across the country, it is helpful to engage in a systematic review of what student learning outcomes data are currently available. Establishing an understanding of the types of data that are currently being collected will create a baseline from which the institution can make decisions about what additional forms of evidence would be useful and if the data currently being collected are identifying actionable next steps.

In addition to summarizing the types of available data, it is important to ask a number of critical questions regarding the usefulness of these data. First, *in the context of the aspirations the institution has for its students and the claims the institution wishes to make about how fully these aspirations are being realized, how directly relevant are these data?* For example, if the institution wishes to demonstrate that it is among the top 25 percent of a set of self-defined peer institutions or similar

institutions within the state/region, then student learning outcomes data that are generated from a locally developed measure may not allow such claims to be made relative to other institutions. In the case of comparisons across institutions, it is clearly essential that the same measures be used at each institution, regardless of whether the measures are developed locally or are commercially available standardized tests. The critical element is common data — to compare apples to apples, as the saying goes. Similarly, if the institution wants to know whether its students are improving in measures of critical thinking, for example, then it is essential the same measures be used each year, that data collection occur on a regular basis, and the statistics provided in reports are comparable.

*The second question that must be asked is whether the data that are available are appropriate for the intended uses.* Some of the primary issues here are the level or unit of analysis, the sampling approach used to collect the data (e.g., whether a random sample of all students was used), and whether the student learning outcomes measures are reliable at the intended level of analysis. Let us consider one of these issues in more detail to illustrate the importance of asking these questions.

Imagine that Institution A is interested in improving the critical-thinking skills of its students. As a result, a decision is made to administer a commercially available assessment that includes critical thinking as one of the skill areas assessed. The institution decides to administer this measure to, say, 200 freshman and 200 seniors. Although the results from this assessment exercise could be used to draw some conclusions about the institution as a whole, it may be difficult (if not impossible) to use the results from this effort to make inferences about the effectiveness of courses that are intended to foster critical-thinking skills. To make statements about these courses, a different sampling plan would be required and attention would need to be paid, for example, to how many students from each course were included in the study.

As this hypothetical example illustrates, it is essential that the relative value or utility of student learning outcomes data be viewed from multiple perspectives. Two critical contexts are identified in Step 1, namely (a) the set of aspirations the institution has articulated and (b) the original purposes for which the data were collected. It may well be that the data do provide important and useful information for the original intention. If this intention/goal is still important, then it is perfectly appropriate and reasonable to continue collecting these data. However, just because these data are available does not necessarily mean that they are appropriate for assessing other desired student learning outcomes or fulfilling other intended uses.

Another set of questions that should be asked regarding the data that are currently being collected is whether the data will reasonably lead to specific actions and if, in fact, these actions are taken. For example, if the data are assessments of graduating students' perceptions of their field of study, have any of the institution's data reports led to changes in curriculum, academic and career advising, or research experiences to address issues identified in the surveys? If the reports are prepared at the department or program level and then forwarded to the appropriate academic officer, are there examples of these reports producing changes at the college, school, or department level?<sup>1</sup> If one looks over a span of several years, is it possible to relate the findings from a report in one year to follow-up actions and results obtained in other years?

### **3. Assessment Augmentation: What Additional Evidence Is Needed?**

It is very likely that the results of Step 2 will indicate gaps between what the institution would like to be able to claim about student learning and what claims the currently available data can support. Because resources are limited in every organization, adding new assessments is almost certain to raise questions about the resources required to support this new initiative, or whether it is even a

---

<sup>1</sup> It is not uncommon on many campuses to have some ongoing data-collection activities that produce annual reports, but generate few, if any, actions. It is reasonable to assume that this pattern of generating reports without associated actions has led to a great deal of faculty skepticism about the utility of student learning outcomes data.

good investment of the limited resources available. These political and fiscal realities will require institutional leaders to find creative ways to respond to these and other challenges. As part of the academic culture, presidents, provosts, vice presidents, deans, and other institutional leaders are regularly asked, in essence, to justify their decisions to invest in some areas and to disinvest in other areas. In the case of assessing student learning outcomes, Steps 1 and 2 advocated here can provide these leaders with important answers to these types of questions. Here are some examples:

The first two steps in this process will create a set of targets for the student learning outcomes system. They will also provide a reality check for how closely the current assessments used within the institution — and the actions that result from the collection of this data — come to addressing key institutional objectives for student success, institutional efficiency, and advancing learning. If the institution has fully engaged in Steps 1 and 2, the results of these efforts will provide a gap analysis between what the institution would like to have as a system for providing claims about student achievement and what is currently possible.

By critically evaluating the currently available student learning outcomes assessment data, it should become clear how the currently available data and the institutional practices for using these data are aligned with the goals identified in Step 1. For example, it is not unreasonable to expect that some of the current assessments will be found to be lacking when viewed in terms of the quality of data needed to address critical institutional issues. Similarly, some assessments or reports that have been used for some time may be found to be out of alignment with current needs or new objectives.

One positive outcome of these analyses is likely to be the realization that some — possibly many — current assessment efforts are providing useful, actionable data, but other efforts are falling short. In the case of the latter, unsuccessful efforts, their retirement can free up resources (faculty and staff time, money, and opportunities for students to participate in assessment efforts) that can be redirected to support new assessment efforts.

Another positive benefit of Step 2 is that institutional leaders will gain a realistic sense of the resources that are being invested in overall assessment efforts. It is likely that this will reveal areas in which there is a duplication of efforts (e.g., separate departments individually undertaking statistical analyses), lack of consistency (e.g., new locally developed assessments being used each year), and other resource drains. Collecting comprehensive information about all ongoing assessment efforts will make it possible to identify areas where efficiencies can be gained (e.g., increasing staff in institutional research may allow for more effective data analysis, report generation, and ongoing monitoring of effectiveness).

#### **4. Refining the Assessment System: Introducing New Assessments, Continuing Valuable Existing Measures**

At this part of the process, an institution should be able to conceptualize its unique student learning outcomes framework. Just as a car's wheels must be realigned periodically, an assessment program may require a similar realignment or adjustment to make sure the full array of measurement tools available is being used to measure the desired student learning outcomes articulated in Step 1. We acknowledge this may be the most difficult step in the process because it may be necessary to retire some assessments and develop or acquire others.

This process provides an opportunity for institutions to consider the mix of internally and externally produced assessments that might be used to measure various student learning outcomes. Institutions may wish to consult COE II for information on 12 currently, commercially available student learning assessments.

## **5. Learning from Our Efforts: What Do the Results from Our Assessment System Tell Us Regarding Our Aspirations for Student Learning?**

Steps 1-4 outline a systematic approach to identifying institutional needs, reviewing existing assessments, and developing a decision-making process to determine how to augment and refine current assessment efforts. These steps set the stage for Step 5, learning from our efforts, which involves the process of analyzing the output from the overall institutional assessment activity.

Step 5 is best viewed as an ongoing, iterative, and cascading process. It is ongoing for two reasons. First, each institution will presumably have some assessment that occurs on a semester, quarter, and/or annual basis. Second, institutional assessment usually never proceeds in a lock-step manner across the campus, with each assessment beginning and ending on a common date. In reality, assessment activities are dispersed over time, and there is considerable overlap in data collection, analysis, review, and reporting. Even with this variability in the schedules of individual assessment efforts, however, the academic calendar does impose some structure on when students are available for testing, when data collection is likely to be completed, and when reports are due.

This combination of schedule variability of individual assessment efforts and the structure imposed by academic calendars creates a situation in which the workload for assessment (in all aspects) is not spread evenly, but rather concentrated in a variety of activities taking place at certain times of the year. For example, assessments of student learning outcomes tend to be administered at the end of the quarter, semester, or year (for traditional courses), and reports are usually generated near the end of the academic year. Unfortunately, the peak times for assessment efforts also often coincide with periods of other peak activities that come at the end of academic terms and years.

This overlap of demands for instructional and assessment efforts creates a situation in which peak workload for faculty and students occurs at a time when the academic and instructional resources of an institution are most concentrated on learning and the assessment of course learning. On many campuses, this creates a tension that can work counter to systematic efforts to collect student learning outcomes data that extend beyond course grades and evaluations. Faculty in these situations are often asked to concentrate on finishing their teaching, grading term papers, writing and grading final exams, and writing letters of recommendation, while also being pressed to conduct important student learning outcomes assessment activities.

There are a number of possible solutions to these and other institutional challenges to creating and sustaining an ongoing and successful student learning outcomes assessment system. We will review several of these here. There are two important points to note before considering how to address these tensions. First, institutional leaders need to recognize these tensions and overlapping workload demands. Second, the institutional leadership needs to send a clear signal, typically some level of resource allocation, that assessment is a high priority. Otherwise, individual faculty members faced with the decision to focus on either their courses and their students, or the demands of an institutionwide assessment effort, will likely choose the former. This does not mean that faculty will resist assessment. It simply means the quality and scope of the assessments that are completed will reflect the many rational decisions that busy faculty make every day. Institutional leaders must recognize that faculty have tremendous dedication and commitment to fostering effective learning, but they also face multiple demands on their time.

Keeping this reality in mind (i.e., that faculty are committed to do the best job they can, but also react to the risks and benefits attendant to decisions of how and where they invest their time and energy), what approaches can be used to ensure that an institution not only has a well-defined set of goals and assessment measures in place, but also adequate attention and resources paid to analyzing the data that result from the assessment efforts?

One key issue institutional leaders must address is the availability of appropriate resources, both human and material. Each institution, depending on its organizational structure and institutional type, will typically have designated individuals or, more commonly, a committee/task force or a unit (institutional research, for example), to undertake initial data collection and basic analysis. Again, depending on internal variables, those undertaking the analysis will probably include faculty and administrators, but may also include other stakeholders (students, for example). It is important for the institution to review whether the current resource mix and set of organizational practices is working effectively, and, more importantly, whether it is likely to work effectively in the future given the newly determined configuration and scope of assessment efforts.

One time-honored tradition in higher education for dealing with new problems and challenges is to create a task force or committee to study the issue and make recommendations. These task forces or committees also often evolve into the unit or committee that becomes responsible for implementing the recommendations and solving the problem. This is relevant in the current context because, more than likely, there already was one or more committees in existence to deal with Steps 1-4. Given the expertise and stature of faculty in higher education, it is also likely that these committees are largely or entirely populated by faculty, with some inclusion of staff and students.

Given the many demands on faculty time (some of which lead to the tension between teaching and other course-level activities and institutional assessments noted earlier), it may benefit the institution to consider an alternative to the traditional committee solution in this context. One possibility, and one that we believe has considerable merit, is to carefully and systematically decide which aspects of assessment data analysis should be handled by faculty and which aspects should be performed by others (e.g., administrative staff, institutional researchers, or experts in assessment and statistical analysts).

Faculty are hired and promoted for their expertise as scholars and teachers. Generally speaking, faculty are not trained to be — nor, in our opinion should they be expected to become — experts in assessment, statistics, and report generation. Rather than assume that faculty should be or become experts in these areas, it may be more appropriate to expect faculty to focus on teaching and scholarship while the institution provides other resources to deal with the demands of assessment. Institutions might consider having institutional experts at their campuses to oversee assessment tasks or possibly work with other institutions to leverage campus-based expertise. We would note that some institutions already utilize surveys and tests from third-party vendors and consortia, while other institutions procure third-party services for test creation (including, for example, validation and analysis of statistical reliability). This approach to making effective use of critical resources (in this case, faculty time) may be extended successfully to other aspects of the assessment effort — especially if the institution has reviewed all of the costs associated with assessment in Step 2.

Let us return now to the key aspects of Step 5. The first element in the process will likely entail a reiteration of the institutional aspirations for student learning and matching selected internal and external measures to those aspirations. Each data set should be analyzed in light of the appropriate aspirational intent. As noted above, if the intent is to use one or more data sets to compare the institution with its peers, it is necessary for the results to be comparable. If the intent is to judge student learning from year to year, then the data collection must occur over time to establish longitudinal trends. Once the data have been clearly linked to the appropriate aspirations, they must be analyzed and evaluated for relevance and validity (see the Validity and Reliability section on page 8).

Measuring the degree to which aspirations for student learning have been successfully achieved occurs at this point. Goals were, undoubtedly, established at the outset of the process, or

expectations of anticipated results may have been created by stakeholders. Now, however, with data in hand, shortfalls between goals and expectations can be identified and successes can be highlighted. In either case, or more likely with the data revealing a combination of successes and shortfalls, the institution can move forward and make the necessary changes and adjustments.

## **6. What Institutional Changes Need to Be Made to Address Learning Shortfalls and Ensure Continued Success?**

The shortfalls and successes in student learning brought into sharper focus by the student learning outcomes data analysis put the college or university in position to determine and implement next steps.

Among the steps that should be considered are:

- A. Communicate and share the results of the data analysis. Communicate to appropriate stakeholders (faculty, students, administrators, the community, and others) the institutional aspirations that were determined to be important and measurable, the tools utilized to measure the aspirations, and the results of the measurements.
- B. Determine, using the internal decision-making processes of the institution, the meaning of the successes and the shortfalls. Identify the steps necessary to address the deficiencies (to move the institution further toward actualization of the aspiration) and to support achievements (to continue the forward momentum for the aspirations attained).
- C. Decisions about the next set of actions, made using the internal decision-making processes and mechanisms (for example, shared governance), should inform budgetary considerations for the ensuing budget cycle. The institution, understanding the importance of certain aspirations for student learning, and identifying its strengths and weaknesses through well-considered measurements, should appropriate the resources needed to address its weaknesses and bolster its strengths. These budgetary decisions will send a signal to the institutional community (students, faculty, administration, governing body, and local and state government) about the institutional leadership's commitment to achieving desired student learning outcomes and employing appropriate measures to quantify the outcomes.
- D. The process, as briefly outlined in A to C, should empower the institution to use its decision-making processes to return to an accountability system model. That is, the institution should recommit to using the results of appropriate assessment instruments to measure achievement of institutional aspirations. The just-concluded cycle of establishing aspirations, identifying appropriate measures, collecting and analyzing the data, and making meaningful decisions based on those data, should coincide with the beginning of the next cycle, but with the institution in a better position to reconsider the initial aspirations and, if necessary, restate them based on the knowledge gained during each step of the cycle.
- E. Finally, the institutional leadership should commit to utilizing the most appropriate tools for the next round of assessment so it can determine whether the changes brought about by the decisions made in Step C have yielded the desired results.

## **7. Ensuring that a Culture of Evidence Is Created Within the Institution: Continuing the Effort Over Time and Expanding to New Areas of Interest**

When it has completed one cycle of the Student Learning Outcomes Model and recommitted to another round, the institution has, in effect, begun the process of institutionalizing the model. By deciding to restate aspirations and rework (if necessary) the tools utilized to acquire relevant data, the institution has created a "Culture of Evidence." The institution, through its commitment to an ongoing process, will also be making a commitment to allocating the appropriate resources (both human and capital), establishing the internal structure, and creating a mechanism for decision making and information sharing.

This dialectic model is consistent with, for example, the “continuous quality improvement model,” with the Southern Association of Colleges and Schools’ Quality Enhancement Plan (QEP) and with the North Central Association’s Academic Quality Improvement Program (AQIP). In making this process iterative, the cycle of improvement becomes a part of the fabric of the institution, not an additional burden or an add-on process, but an integral component of how the institution judges itself and plans for its future.

There are multiple benefits that can be derived from this process. As noted above, accrediting agencies almost uniformly ask for evidence of a process of internal decision making that ties institutional goal setting to data collection to budget and other decision-making processes. This process can become a key element in an institution’s strategic planning and reaccreditation efforts.

As the accountability paradigm is embraced by the higher-education community, this process for assessing student learning outcomes will allow colleges and universities to define their own unique goals and establish the standards on which they wish to be held accountable. In this way, an institution can demonstrate its commitment to accountability and the progress being made in achieving its institutional aspirations to important shareholders (e.g., state officials, in the case of public institutions; boards of trustees; students; parents; and the community).

## CONCLUSION

---

Education is changing rapidly at the start of the 21st century. There is a growing need for institutions to evolve to meet the changing needs of students and to stay abreast of continuous advances in knowledge and technology. The central activity of any college or university, its *raison d’être*, is learning and teaching. By making measurement of student learning an integral component of the institution and tying it to its institutional aspirations, colleges and universities make manifest their commitment to the core mission of student learning.

In this report, we have outlined an approach that we believe will be useful to higher-education leaders as they grapple with the challenges posed by the new accountability paradigm. This approach has a number of benefits that we believe will accrue to those leaders and institutions that adopt this approach.

First, by involving a range of stakeholders in the process and by broadly communicating the results of the process internally and externally, the institution creates a transparent system of accountability. This transparency is consistent with current expectations of higher-education stakeholders. It will enhance the image of the institution in the community and the region.

Second, the broader community of stakeholders (students, faculty, administration, governing body, funding agencies, and government officials) can be assured through this process that institutional resources are being utilized in appropriate and cost-effective ways. As the process works through its cycles, as hard questions are asked, and as activities that are not producing expected results are eliminated or scaled back, a college or university can show through its budgetary actions that it is a good steward of institutional resources.

Third, this approach can support a faculty and student focus on the science of learning and pedagogy as a scholarly activity, especially within specific academic disciplines. Almost all academic discipline societies have a division or council that focuses on teaching and learning (e.g., American Psychological Association’s Division 2–Society for the Teaching of Psychology). Institutions will also gain recognition for creating and advancing best practices that support student achievement.

As a nonprofit organization, ETS puts the needs of students and educators first. Although this report is the last in a planned series of Culture of Evidence papers, ETS will continue its mission of

promoting learning worldwide by remaining on the forefront of assessment research. New research and white papers by some of the world's leading researchers, psychometricians, and assessment experts — leading authorities in educational measurement — are posted regularly on ETS's website, [www.ets.org](http://www.ets.org).

We invite you to learn more about ETS's research and work in measuring student learning outcomes by visiting [www.ets.org/cultureofevidence](http://www.ets.org/cultureofevidence).

## ANNOTATED BIBLIOGRAPHY

---

Adelman, C. (2006). *The toolbox revisited: Paths to degree completion from high school through college*. Washington, DC: U.S. Department of Education.

This report follows a nationally representative cohort of students to determine which aspects of their education contributed to bachelor's degree completion. The study describes seven analytical steps linked chronologically to student pathways through higher education, providing numerous tables to facilitate quick observation and comparison of factors that contribute to degree completion. Five of these — generation of more than 20 credits during the first year, repeal of lax no-penalty withdrawal and no-credit repeat policies, use of summer terms, prompt post-secondary entry, and more rigorous high school curriculums — stand out as positively influencing student bachelor's degree completion rates. The complex pathways reflected in the data also call into question traditional formulas to determine college graduation rates. Themes, highlights, and implications for secondary- and higher-education leaders and policymakers are discussed.

Bok, D. (2006). *Our underachieving colleges: A candid look at how much students learn and why they should be learning more*. Princeton: Princeton University Press.

In this book, former Harvard University President Derek Bok analyzes empirical evidence to assess the shortcomings of postsecondary education in four-year colleges and universities in a candid and balanced fashion. After examining the history of U.S. colleges and the debate about the goals of higher education, Bok extensively reviews the literature on student learning and proposes eight student learning outcomes as central to the educational missions of colleges and universities: communication skills, critical-thinking skills, moral reasoning, citizenship preparation, appreciation for diversity, preparation for a global society, the development of interests, and career preparation. In the final chapter, Bok carefully critiques the potential effectiveness of various policies for promoting measurable improvements in postsecondary student achievement in these areas.

Council of Regional Accrediting Commissions (2004). *Regional accreditation and student learning: Improving institutional practice*. Retrieved July 17, 2007, from the Southern Association of Colleges and Schools Commission on Colleges website: <http://www.sacscoc.org/pdf/handbooks/ImprovingPractice.pdf>.

A more detailed version of the companion guide, *Regional Accreditation and Student Learning: A Guide for Institutions and Evaluators*, this concise and well-referenced booklet explores how institutions can effectively prepare for accreditation by focusing on student learning. It provides a framework for colleges and universities who share a commitment to five principles of institutional quality: the role of student learning in accreditation, documentation of student learning, compilation of evidence, stakeholder involvement, and capacity building. Each section focuses on one of these principles and includes information compiled from experts and suggestions for improving institutional practice. An indexed bibliography and list of criteria for evaluating student learning complete the guide.

Driscoll, A., & Cordero de Noriega, D. (2006). *Taking ownership of accreditation: Assessment processes that promote institutional improvement and faculty engagement*. Sterling, VA: Stylus.

This guide depicts accreditation as an opportunity to engage faculty, students, and administrators with the assessment of student learning outcomes. Determined to go beyond satisfying accreditation requirements, the authors focused accreditation processes at their own institution, California State University, Monterey Bay, on supporting the continuous, lasting improvement of educational practice and student learning. Practical strategies that campuses can adapt or apply as-is to meet their institutional assessment needs are presented, along with accreditation documentation guidelines and first-hand advice for gaining faculty support by emphasizing the importance of assessment data for improving the students' learning, the faculty members' teaching, and the quality of academic courses, departments, and programs.

Ewell, P. (2006). *Making the grade: How boards can ensure academic quality*. Washington, DC: Association of Governing Boards of Colleges and Universities.

In this book, Peter Ewell provides a clear and concise overview of what institutional governing board members need to know, not just about the fiscal health of colleges and universities, but also about their academic well-being. After briefly reviewing new developments in higher education and our global context that have contributed to increased scrutiny of students' academic outcomes, he goes on to describe four major areas of academic quality assessment: student learning outcomes, student retention and graduation, stakeholder satisfaction, and program review. In each section, basic concepts and processes and the questions board members should ask their institutions about them are reviewed, emphasizing interests and responsibilities particular to boards. The author also discusses accreditation as an integral element of academic quality review that can provide guidance for institutional improvement as well as external verification that colleges and universities are fulfilling claims made about their students' learning.

Kuh, G. D., Kinzie, J., Schuh, J. H., & Whitt, E. J. (2005). *Assessing conditions to enhance educational effectiveness: The inventory for student engagement and success*. San Francisco: Jossey-Bass.

This book details the Inventory for Student Engagement and Success (ISES), a framework for systematic and comprehensive institutional self-assessment. Part one presents a rationale for self-assessment and explores the guiding principles for applying the ISES framework. The second part provides in-depth descriptions for using the ISES framework to assess the culture and conditions of effective colleges, to assess the prevalence of effective practices on campus, and to guide institutional improvement. Included are diagnostic questions focused on six conditions common to high-performing schools (a "lived" mission and philosophy of education, unshakeable focus on student learning, educationally enriching environment, clear pathways to student success, an improvement-oriented culture, and shared responsibility for student success) and the five areas of effective educational practices assessed by the National Survey of Student Engagement (challenging academics, active and collaborative learning, student-faculty interaction, enriching experiences, and supportive environments).

Palomba, C. A., & Banta, T. W. (1999). *Assessment essentials: Planning, implementing, and improving assessment in higher education*. San Francisco: Jossey-Bass.

"Assessment is the systematic collection, review, and use of information about educational programs undertaken for the purpose of improving student learning and development." Starting from this basic premise, *Assessment Essentials* goes on to review the assessment process from articulating definitions and goals to communicating and applying assessment results for campus improvement. The clear, accessible language and wealth of illustrative examples make this text especially useful for faculty members and campus administrators who may be unfamiliar with the campus assessment process. This comprehensive resource covers everything from instrument reliability and validity to involving student affairs professionals with faculty in assessment.

Schechter, E. (2007). *Internet resources for higher education outcomes assessment*. Retrieved July 17, 2007, from North Carolina State University, University Planning & Analysis website: <http://www2.acs.ncsu.edu/UPA/assmt/resource.htm>.

This linked online resource provides frequently updated information about higher education outcomes assessment in general, as well as related to specific tests, major content areas, institutional practices, and accrediting bodies. The section "Assessment Handbooks" provides a linked list of institutional assessment guides, while the "Assessment of Specific Skills or Content" section has information and links to standardized tests and other assessment tools, followed by discipline-specific resources that often include statements of standards and/or links to locally developed measures. Finally, Individual Institutions' Assessment-Related Pages is composed of a lengthy list of colleges and universities involved in outcomes assessment, most of which are briefly annotated and linked to their assessment web pages.

Walvoord, B. E. (2004). *Assessment clear and simple: A practical guide for institutions, departments, and general education*. San Francisco: Jossey-Bass.

Walvoord's book is a concise, step-by-step guide filled with practical advice and straightforward principles of assessment that can be applied and adapted for various disciplines and institution types. With well-segmented sections within each chapter and a liberal dose of bulleted and numbered lists, *Assessment Clear and Simple* serves as a true guidebook, beginning with an overview of concepts, moving into material geared toward stakeholders at different levels within and beyond the institution, and concluding with a discussion of general education and assessment. Annotations of recommended sources for further reading supplement each chapter, and appendices provide samples of rubrics for student evaluations, guidelines for program reviews, assessment reports, and other useful resources.

## REFERENCES

---

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Association of American Colleges and Universities. (2002). *Greater expectations: A new vision for learning as the nation goes to college*. Washington, DC: Author.
- Association of American Colleges and Universities. (2007). *Rising to the challenge: Meaningful assessment of student learning* (Project Overview). Retrieved November 7, 2007, from <http://snipurl.com/1te7u>
- Baer, J. D., Cook, A. L., & Baldi, S. (2006). *The literacy of America's college students*. Washington, DC: American Institutes for Research.
- Basken, P. (2007). A year later: Spellings report still makes ripples. *The Chronicle of Higher Education*, 54(5), A1. Retrieved November 6, 2007, from <http://snipurl.com/1uwtx>
- Bollag, B. (2006). Making an art form of assessment. *The Chronicle of Higher Education*, 53(10), A8-A10. Retrieved November 6, 2007, from MasterFILE Premier database <http://snipurl.com/1uwvq>
- The Business-Higher Education Forum. (2004). *Public Accountability for Student Learning in Higher Education: Issues and Options* (Rep. No. 309696). Washington, DC: American Council on Education.
- Business Roundtable. (2005). *Tapping America's potential: The education for innovative initiative*. Washington, DC: Author.
- Committee on Science, Engineering, and Public Policy. (2007). *Rising above the gathering storm: Energizing and employing America for a brighter economic future*. Retrieved October 13, 2005, from National Academies Press website <http://snipurl.com/1te7y>
- Curriss, C. W., & Lingenfelter, P. E. (2005). A unit-record system of collecting student data would broaden accountability in nontraditional times: Opinion. *The Chronicle of Higher Education*. Retrieved November 6, 2007, from <http://snipurl.com/1v3wh>
- Dwyer, C. A., Millett, C. M., & Payne, D. G. (2006). *A culture of evidence: Postsecondary assessment and learning outcomes*. Princeton, NJ: ETS.
- Education Commission of the States. (1998). *The progress of education reform 1998*, Denver, CO: Author.
- Eubanks, D. (2006). *The problem with standardized assessment: There are other, better ways than high-stakes testing to hold institutions accountable for making good on the promises of higher education*. Retrieved November 5, 2007, from <http://snipurl.com/1te81>
- Frye, R. (1999). Assessment, accountability, and student learning outcomes. *Dialogue*, (2), 1-12. Retrieved November 4, 2007, from Western Washington University, Office of Institutional Assessment and Testing website <http://snipurl.com/1te85>
- Garcia, A. E., & Pacheco, J. M. (1992, March). *A student outcomes model for community colleges: Measuring institutional effectiveness*. Paper presented at the 1992 North Central Association of Colleges and Schools commission. Retrieved November 7, 2007, from the ERIC database <http://snipurl.com/1te88>
- Kirsch, I., Braun, H., Yamamoto, K., & Sum, A. (2007). *America's perfect storm: Three forces changing our nation's future*. Princeton, NJ: ETS.
- Kotter, J. P. (1996). *Leading change*. Boston, MA: Harvard Business School Press.

- Labi, A. (2007). International assessment effort raises concerns among education groups. *The Chronicle of Higher Education*, 54(5), A31-A31. Retrieved November 6, 2007, from Education Research Complete database <http://snipurl.com/1v1op>
- Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher*, 23(2), 13-23.
- Miller, M. (2006). The legitimacy of assessment. *The Chronicle of Higher Education*, 53(5), B24.
- Millett, C. M., Stickler, L. M., Dwyer, C. A., & Payne, D. G. (2007). *A culture of evidence: Critical features of assessments for postsecondary student learning*. Princeton, NJ: ETS.
- Mislevy, R. J., Almond, R. G., & Lukas, J. F. (2003). *A brief introduction to evidence-centered design*. CSE Technical Report ETS Research Rep. No.RR-03-16. Los Angeles, CA: The National Center for Research on Evaluation, Standards, Student Testing (CRESST), Center for Studies in Education, UCLA. Retrieved November 6, 2007, from <http://snipurl.com/1te8d>
- National Association of Independent Colleges and Universities. (1994). *The responsibility of independence: Appropriate accountability through self-regulation*. Final Report of the NAICU Task Force on Appropriate Accountability: Regulations, Accreditation and Assessment. Washington, DC: National Association of Independent Colleges and Universities.
- National Center for Educational Statistics. (2006). *Digest of Education Statistics 2006, table 248*. Retrieved November 7, 2007, from <http://snipurl.com/1te8g>
- Organization for Economic Co-operation and Development. (2007). *Education at a glance 2007: OECD indicators*. Retrieved November 7, 2007, from <http://gsed.wordpress.com>
- Palomba, C. A., & Banta, T. W. (1999). *Assessment essentials: Planning, implementing, and improving assessment in higher education*. San Francisco: Jossey-Bass.
- Presidents' Forum of Excelsior College. (2007). *Transparency by design: Principles of good practice for higher education institutions serving adults at a distance*. Retrieved November 7, 2007, from <http://snipurl.com/1te8j>
- Schagen, I. & Hutchison, D. (2007). Comparisons between PISA and TIMSS: We could be the man with two watches. *Education Journal*, 101, 34-35.
- Snider-Lotz, T. G. (2002). *Designing an evidence-centered assessment program*. Atlanta, GA: Qwiz, Inc.
- State Higher Education Executive Officers. (2005). *Accountability for better results: A national imperative for higher education*. Denver, CO: Author.
- U.S. Department of Education. (2006). *A test of leadership: Charting the future of U.S. higher education*. Washington, DC: Author.





Rosedale Road  
MS 30-N  
Princeton, NJ 08541

STANDARD MAIL  
US POSTAGE  
PAID  
EDUCATIONAL  
TESTING SERVICE

746122

