National Institute for Learning Outcomes Assessment

November 2009

Assessment, Accountability, and Improvement:

Revisiting the Tension

Peter T. Ewell

Foreword by George D. Kuh



Occasional Paper #1

learningoutcomesassessment.org

About the Author

Dr. Peter Ewell

Dr. Ewell is the Vice President at the National Center for Higher Education Management Systems (NCHEMS), a research and development center founded to improve the management effectiveness of colleges and universities. A member of the staff since 1981, Dr. Ewell's work focuses on assessing institutional effectiveness and the outcomes of college, and involves both research and direct consulting with institutions and state systems on collecting and using assessment information in planning, evaluation, and budgeting. He has directed many projects on this topic, including initiatives funded by the W. K. Kellogg Foundation, the National Institute for Education, the Consortium for the Advancement of Private Higher Education, and The Pew Charitable Trusts. In addition, he has consulted with over 375 colleges and universities and more than thirty state or national governments internationally on topics including assessment, program review, enrollment management, and student retention. Dr. Ewell has authored seven books and numerous articles on the topic of improving undergraduate instruction through the assessment of student outcomes.

A graduate of Haverford College, he received his Ph.D. in Political Science from Yale University in 1976 and was on the faculty of the University of Chicago.

No one is more qualified than Peter Ewell to author the inaugural paper in the NILOA Occasional Paper Series.

—George Kuh

Contents

Abstract 3

Foreword 4

Assessment, Accountability, and Improvement 5

What Has Changed? 5
The Tension: Then and Now 7

Appropriate Domains of Accountability: Who Owes What to Whom? 9

Managing the Tension: Some Principles of Response 14
Concluding Thoughts 20

References 21

NILOA 22

National Advisory Panel 22

NILOA Mission 22

Occasional Paper Description 22

About NILOA 23

NILOA Staff 23

NILOA Sponsors 23

The ideas and information contained in this publication are those of the authors and do not necessarily reflect the views of Carnegie Corporation of New York, Lumina Foundation for Education, or The Teagle Foundation.



Abstract

Assessment, Accountability, and Improvement: Revisiting the Tension

Many of the same tensions that characterized the accountability and improvement purposes of student learning outcomes assessment when the assessment movement began in the mid-1980s still exist today. In this paper I examine these tensions and how they can be managed, if not completely resolved. First, I outline the major relevant changes affecting the assessment movement that have occurred in higher education over the past two decades. These include the perceived legitimacy of assessment today, the demand by policymakers for better and more transparent information about student and institutional performance, the press by accreditors on institutions to collect and use student learning outcomes data, and the availability of more and better assessment instruments and approaches.

Then, I describe and analyze the conceptual incongruities between the accountability and improvement assessment paradigms. Adopting either of these two perspectives affects institutional choices about what and how to assess, how to organize assessment tasks and strategies, and how to communicate assessment results. As with all ideal types, the differences between these two contrasting opposing paradigms of assessment are exaggerated, and rarely does an existing assessment approach fully conform to either one. The next section discusses the major external players in higher education that have stimulated institutions to engage in assessment and the kinds of information about performance on which they do or should focus. The groups include state government agencies, the federal government, regional and specialized accreditors, and the public interest represented by consumer demand for information and third party judgments (e.g., rankings) about institutional performance. I close by discussing four principles to help guide institutions in successfully dealing with the tensions between improvement and accountability and the sometime competing interests of internal and external stakeholders: (1) respond visibly to domains of legitimate external concern; (2) show action on the results of assessment; (3) emphasize assessment at the major transition points in a college career; and (4) embed assessment in the regular curriculum.

Despite adhering to these principles and using other emerging promising practices, some elements of the accountability-improvement tension may be difficult to completely resolve. Nevertheless, because the stakes associated with institutional performance are so much higher for policy makers today, it is imperative that we make much more progress in collecting and using assessment results to improve and in communicating what we are doing more effectively to external audiences.



Foreword

Assessments of what students learn during college are typically used for one of two purposes – improvement or accountability. In the former, faculty members and other institutional personnel gather evidence about how well students are attaining intended course, program, or institution outcomes, and then use this information to improve student performance by modifying pedagogical approaches as well as institutional policies and practices. In this sense, assessment for improvement is essentially an internal matter. In contrast, assessment data collected for the purpose of accountability are used primarily to demonstrate that the institution is using its resources appropriately to help students develop the knowledge, skills, competencies, and dispositions required to function effectively in the 21st century. The information is typically intended for external audiences.

For reasons carefully outlined in this NILOA Occasional Paper, since the early days of the "assessment movement" in the US, these two purposes of outcomes assessment have not rested comfortably together. Indeed, the author of this paper, Peter Ewell, once characterized the relationship between assessment for improvement and assessment for accountability as a "contradiction." During an animated discussion of the NILOA National Advisory Panel in December, 2008, two points of consensus emerged related to these purposes of assessment. First, too many faculty, staff, and policy makers view improvement and accountability to be trains running on parallel tracks, without the promise of converging to an intersection that provides enriched answers to questions related to both purposes. The second point of consensus was that a fresh, informed, thoughtful analysis of the relationships between the improvement-accountability purposes could bring the two tracks closer together, if not resolve all the tensions their uses present.

No one is more qualified than Peter Ewell to author the inaugural paper in the NILOA Occasional Paper Series. And no one is better prepared than Peter to flesh out the factors and perspectives that contribute to what for several decades has been interpreted at best as an uneasy relationship between the twin purposes of assessment and at worst the contradiction that fuels suspicion among skeptical faculty about whether assessment can be a lever for improved student learning and, therefore, worthy of their time and energy. He has written multiple papers on assessment, contributed to numerous policy and professional meeting discussions including participating on the working group that produced the American Association for Higher Education (AAHE) Principles of Good Practice for Assessing Student Learning in 1992, and presently sits on several advisory boards that are guiding initiatives that bear directly and indirectly on the assessment agenda. In addition to his role as vice president of the National Center for Higher Education Management Systems, Ewell is a NILOA Senior Scholar and was an original member of NILOA's National Advisory Panel.

In this paper, Ewell draws on his quarter century of experience as a participant-observer and chronicler of the assessment movement to summarize what has changed and what has not over the past two decades in terms of the assessment of student learning and the shifting expectations and demands of policy makers, accreditors, higher education leaders, and government officials about student and institutional performance. After delineating how various kinds of information can and should be used for improvement and accountability, he points to ways that institutions can productively manage the persistent tensions associated with improvement and accountability as faculty and staff members do the important work of documenting, reporting, and using what students have learned and can do as a result of their college experience.

George D. Kuh
Chancellor's Professor and Director
NILOA and Indiana University Center for Postsecondary Research



Assessment, Accountability, and Improvement: Revisiting the Tension

Peter T. Ewell

My files have a paper, *Assessment, Accountability, and Improvement: Managing the Contradiction* (Ewell, 1987), that I prepared for the Third National Assessment Forum of the American Association for Higher Education (AAHE) in 1987. Twenty-two years later, that organization and the conference for which it was prepared no longer exist, but the points I made then seem alarmingly current. On the one hand, this is depressing. After almost a quarter century and a lot of disputation and spilled ink on assessment, we might have expected more progress—or at least more nuance and sophistication when the topic is discussed. On the other hand, the apparent timelessness of these issues suggests they are hard and important—factors prompting their re-examination. This paper, commissioned by the National Institute for Learning Outcomes Assessment (NILOA), at the invitation of George Kuh and Stan Ikenberry, is my attempt to do so.

This paper follows the general logic of the earlier work by first presenting the conceptual tensions between assessment for accountability and assessment for improvement. Adopting either one of these two perspectives will decisively influence institutional choices about what and how to assess, how to organize assessment, and how to communicate assessment results. My current reflections on these conceptual issues—drawn from several papers I have published on these topics in the intervening years—constitute the second main section of the present paper. The third section discusses the major external players in higher education that have stimulated institutions to engage in assessment by looking at their basic interests and the kinds of information about performance that they are, or should be, looking for. In the fourth section, I offer advice for how institutions can navigate the sometimes competing tensions between improvement and accountability. Before doing any of this, though, I briefly outline how the contours of the higher education landscape have changed since the late 1980s in ways that will affect the topic.

Adopting either one of these two perspectives will decisively influence institutional choices about what and how to assess, how to organize assessment, and how to communicate assessment results.

What Has Changed?

In 1987, the so-called "assessment movement" in U.S. higher education was less than five years old. It had in part been stimulated by a combination of curriculum reform reports that called for greater curricular coherence, the use of powerful pedagogies known to be associated with high learning gains,[1] and knowledge about student outcomes and experiences (Ewell, 2002). But it had also been given impetus by the growing interests of state governments in using newly available tests and surveys to demonstrate return on investment. These were the roots, more than 25 years ago, of the tension between assessment's use for accountability and its use for internal improvement—a tension that has dogged assessment ever since. Most of the central elements of this tension remain with us today. But the passage of a quarter century has also changed much of the environment within which these elements and the tension play out. Among the most important of these changes have been the following:

[1] Summarized that year as the famous Seven Principles for Good Practice in Undergraduate Learning (Chickering & Gamson,1987).

- ***Perceived Legitimacy.*** Although assessment probably remains distasteful for many faculty members, it is angrily rejected far less frequently than it was 20 years ago. The majority of academics now realize that engaging in assessment has become a condition of doing business for colleges and universities because of accreditation requirements and the need to show results to taxpayers and potential customers. All of them may not want to engage in assessment themselves, but they are willing to accord the activity a legitimacy that was not forthcoming two decades ago. This is no small thing, because it means that the sharpest tension embedded in assessment's purposes is less apparent today. The question has become more about what kinds of assessment to engage in and under whose control than about whether or not to engage in it at all. Meanwhile, a sizeable minority of faculty have wholeheartedly embraced assessment as useful in improving undergraduate instruction—a constituency of adherents that did not exist a quarter century ago. Many of these faculty are in disciplines where, because of programmatic accreditation requirements, assessment practices are common—disciplines like teacher education, engineering, and the health professions. Buy-in among these groups has also been helped by the growing availability of evidence-gathering approaches like classroom research and surveys of student engagement that provide immediately useful feedback about instructional practices.

- ***New Policy Centrality of Higher Education***. The broader policy environment for higher education has also shifted substantially since 1987. Arguably, at that time the perceived need among policy makers to raise average citizens' levels of educational attainment was less urgent than in today's world, where the competitive position of the U.S. with respect to educational attainment is beginning to slip (Organization for Economic Cooperation and Development, 2008). As reflected in President Obama's ambitious goal of 60 percent of U.S. young adults with a college credential, higher education is seen by virtually all opinion leaders as more important than ever before. And this perceived centrality is bipartisan and deep. The sentiments expressed in the report of the "Spellings Commission" (U.S. Department of Education, 2006) are shared widely across the policy and business communities and will continue to press higher education for accountability in ways that were inconceivable two decades ago. The central leitmotifs of this new accountability environment are transparency and learning outcomes. Colleges and universities are being asked to disclose more and more about academic results and are responding in kind. Most now realize that it will be impossible to sit out the latest round of pressure for accountability with the hope that it will eventually go away. The current choice is between proactively taking responsibility for demonstrating accountability on the academy's own terms or passively having requirements dictated from the outside with little or no control.

- ***From States to Accreditors as External Stimuli.*** Two decades ago, the principal actors external to colleges and universities requiring attention to assessment were state governments. Proactive mandates in such states as Colorado, Missouri, New Jersey, South Dakota, Tennessee, Virginia, and Washington unfolded in a largely benign fiscal environment. After a few rough bouts with standardized testing, state leaders were persuaded that assessment approaches allowing institutions to set their own goals for student learning and to develop institution-specific (therefore non-comparable) methods for gathering evidence of their achievement could both aid improvement and should be sufficient to discharge accountability—so long as institutions acted in good faith to collect the evidence and actually use it. Meanwhile, regional accreditors had entered the assessment arena, and by the early 1990s all of them had standards requiring "institutional effectiveness" or "assessment" that looked a good deal like the state mandates of the prior decade (Ewell, 2002).[2] By this time, most states had

The most important changes in the assessment movement have been the following:

- *Perceived Legitimacy*
- *New Policy Centraility of Higher Education*
- *From States to Accreditors as External Stimuli*
- *Assessment Technology*

[2] A similar progression occurred with most of the specialized accreditors over the same time period that produced even more development in the fields affected—engineering, education, business, and the health professions.

stopped enforcing their mandates because of both budget constraints and a movement of attention toward performance funding and higher degree-completion rates. This shift from state governments to regional accreditors as the locus of external stimulus for assessment has had important implications. One is that independent institutions are as affected as public institutions. Because of its peculiar position as an "accountability" actor jointly owned by the academy and the federal government, moreover, accreditation can buffer the assessment-for-accountability relationship in ways not available to governmental regulation and can simultaneously promote improvement (Ewell, 2008b).

- ***Assessment Technology.*** In 1987, only a handful of instruments were available designed specifically to assess the effectiveness of undergraduate education. As a result, most institutions engaging in assessment used a mixture of surveys (both commercial and home-grown) and a collection of cognitive examinations like the ACT Assessment or the Graduate Record Examination that were designed for something else. Those purpose-built instruments that were available (with the exception of the pioneering long form of the ACT College Outcomes Measures Project), moreover, were standardized multiple-choice tests deemed inappropriately narrow and under-nuanced by most college faculty. Mandated use of such "philistine" approaches was thus doubly unpalatable. Today, in contrast, campuses engaging in assessment have a range of choices generally more acceptable to faculty. These include "authentic" or task-based assessments using constructed-response formats like the Collegiate Learning Assessment (CLA), approaches using actual student work products or artifacts (increasingly compiled in electronic portfolios), "embedded" assignments in regular courses periodically graded by rubric, and well-established surveys like the National Survey of Student Engagement (NSSE) that can inform instructional improvement directly. Using such approaches, of course, cannot in itself render externally mandated assessment comfortable for faculty, but it can mitigate the discomforts somewhat and allow faculty to learn something useful from assessment.

The changes in the landscape of higher education, while important, do not alter the fundamental terms of the "contradiction" I presented in 1987 between assessment's use for accountability and its use for improvement. The changes do serve, however, to complicate and modulate the relationships among the range of potential accountability actors. Most of these developments make the dichotomy itself a lot less stark. What was once a "contradiction" has in many ways become an embedded "tension"—never wholly resolvable, but to some extent responsive to institutional attention. On the other hand, because the stakes associated with higher education are so much higher for policy makers today, aggressive action on the accountability agenda is more likely and a proactive response on the part of the academy is more urgent.

Because the stakes associated with higher education are so much higher for policy makers today, aggressive action on the accountability agenda is more likely and a proactive response on the part of the academy is more urgent.

The Tension: Then and Now

The conflicting imperatives of accountability and improvement that formed the basis of my argument two decades ago remain substantially intact. Accountability requires the entity held accountable to demonstrate, with evidence, conformity with an established standard of process or outcome. The associated incentive for that entity is to look as good as possible, regardless of the underlying performance. Improvement, in turn, entails an opposite set of incentives. Deficiencies in performance must be faithfully detected and reported so they can be acted upon. Indeed, discovering deficiencies is one of the major objectives of assessment for improvement. The state mandates that constituted the assessment-as-accountability pole of the tension 20 years ago, in the main, had effects that were similar to assessment's most recent embodiment in the recommendations of the Spellings Commission: an institutional response oriented largely toward compliance and one typified by noticeable

faculty disengagement. The central conclusion of this dynamic, then as now, is that when institutions are presented with an intervention that is claimed to embody both accountability and improvement, accountability wins.

Twenty years ago there also was a good deal of complexity surrounding the tension that belied straightforward polar opposition. To begin with, virtually all of the states mandating assessment professed to be in the business of "improvement." Most eschewed standardized testing for this reason, rightly believing that the relatively few and largely acontextual numbers that testing generates would be ill-suited for informing instructional change. State assessment mandates also typically resulted in a bimodal distribution of institutional response. Most institutions complied, wrote plans, and "did assessment." But not much of what they did was connected to their regular planning, budgeting, and internal review activities. A few of them, though, embraced the process wholeheartedly, using the mandate to mobilize initial action but carrying assessment through multiple cycles unasked and driving it down to the lowest levels of the organization. So a lot of the tension between the polar opposites that I argued more than 20 years ago was more nuanced even then. How this dynamic plays out can also be affected by where when one sits in an organizational hierarchy—for example, the "accountability" dynamic can be as applicable to the relationship between a history professor and a dean as to that between a university president and a governor.

When institutions are presented with an intervention that is claimed to embody both accountability and improvement, accountability wins.

I have fleshed out elsewhere these different roles of assessment in two "paradigms" (Ewell, 2008b). These contrasting "ideal types" are worth describing again here, because the relationships between them can get complicated. The first, evolving from the "institution-centered" approach of the mid-1980s, can best be described as the "Improvement Paradigm." The second, derived from the early state mandates, can best be labeled the "Accountability Paradigm." Table 1 contrasts these two "ideal types" along a number of dimensions.

Table 1

Two Paradigms of Assessment

	Assessment for Improvement Paradigm	Assessment for Accountability Paradigm
Strategic Dimensions		
Intent	Formative (Improvement)	Summative (Judgment)
Stance	Internal	External
Predominant Ethos	Engagement	Compliance
Application Choices		
Instrumentation	Multiple/Triangulation	Standardized
Nature of Evidence	Quantitative and Qualitative	Quantitative
Reference Points	Over Time, Comparative, Established Goal	Comparative or Fixed Standard
Communication of Results	Multiple Internal Channels and Media	Public Communication
Uses of Results	Multiple Feedback Loops	Reporting

These contrasting conceptual pictures of assessment can be distinguished around two sets of descriptors—the first, "Strategic Dimensions," addresses purpose and strategy while the second, "Application Choices," addresses method and implementation. "Intent," the first heading under "Strategic Dimensions," describes what the process is designed to accomplish. Within the "Improvement Paradigm," the intent is to use the resulting information to enhance teaching and learning. Within the "Accountability Paradigm," in contrast, the intent is to use information to demonstrate to policy makers and the public that the enterprise they fund is effective and worth supporting. The second heading, "Stance," describes the standpoint from which evidence-gathering and application is undertaken. Within the "Improvement Paradigm," evidence is gathered and judged directly by practitioners (faculty and administrators) acting within the parameters of the teaching and learning process. Within the "Accountability Paradigm," evidence is gathered and judged from an external "objective" standpoint. The final heading, "Predominant Ethos," addresses the prevailing mentality surrounding assessment. Within the "Improvement Paradigm," the predominant ethos is a posture of engagement seeking continuous improvement and a "culture of evidence." Within the "Accountability Paradigm," the predominant ethos is a posture of institutional compliance, or at least the appearance of it.

These basic distinctions of purpose and strategy imply some important further distinctions in the way the assessment process is enacted. With respect to methods for gathering evidence, the "Improvement Paradigm" can embrace many kinds of evidence-gathering including standardized and faculty-designed examinations, capstone projects, demonstrations, portfolios, and specially designed assignments embedded in regular courses. But because its principal purpose is comparison, the "Accountability Paradigm" places greater reliance on standardized examinations and surveys. Similarly, assessment evidence under the "Improvement Paradigm" can be both quantitative and qualitative, while evidence under the "Accountability Paradigm" is almost exclusively quantitative. In parallel, the reference points used for judging results under the "Improvement Paradigm" can include tracking progress over time or against established institutional goals. It can also involve comparing results across units or for different kinds of students. For the "Accountability Paradigm," however, it is centered on comparisons across institutions or programs, or against fixed standards of performance. Multiple channels of communication are used to disseminate assessment results to a variety of internal constituents under the "Improvement Paradigm," while the "Accountability Paradigm" relies instead on transparent public reporting. Finally, assessment results are used primarily to guide intervention through the establishment of multiple feedback loops under the "Improvement Paradigm," while results are used in the "Accountability Paradigm" primarily to assure external stakeholders of the effectiveness of their investments through comparative reporting.

As is the case for all ideal types, the differences between these two opposing paradigms of assessment are exaggerated, and almost no existing assessment approach fully conforms to either of them. But this conceptual tool does serve to lay out some fundamental tensions embedded in the different ways of implementing assessment that have been with us from the beginning.

Appropriate Domains of Accountability: Who Owes What to Whom?

Assessment has been stimulated by many external actors over the last quarter century including states, the federal government, accrediting organizations, and various third-party organizations—each with its own specific interests in evidence on institutional and program performance. These differing interests, some of which have changed over time, must be distinguished and clearly

Assessment has been stimulated by many external actors over the last quarter century including states, the federal government, accrediting organizations, and various third-party organizations—each with its own specific interests in evidence on institutional and program performance.

understood as a prerequisite for developing a proper response by the academy.

The States. State governments, rather than the federal government, have primacy in the U.S. because they are the units of organization and funding for the higher education enterprise. State treasuries provide direct support for public colleges and universities and frequently provide funding indirectly to independent institutions through student aid programs. States also grant institutions of all kinds—public, independent, and proprietary—permission to operate in the first place through registration or licensure. These basic functions remain today just as they were in place a quarter century ago, but two things have changed. First, there is a whole lot less funding for anybody, due to increasingly tight state budget constraints. Second, states have begun to embrace a broader and more integrated approach to higher education policy that attempts to create a conscious "public agenda" for higher education (Ewell & Jones, 2006). Both of these have affected how states think about accountability and assessment.

With regard to money, the conventional wisdom has it that accountability goes up as money gets tight. This appears not to be panning out with respect to learning outcomes assessment, however.

With regard to money, the conventional wisdom has it that accountability goes up as money gets tight. This appears not to be panning out with respect to learning outcomes assessment, however. Most of the mandates enacted by more than two thirds of the states had by 1990 either fallen by the wayside entirely or were no longer enforced, and the handful of prominent statewide testing programs that commentators like me used to point to have mostly disappeared. Florida and Texas gave up their "gateway" testing programs a few years ago, and states like Arkansas, Colorado, Utah, and Wisconsin that for at least a couple of years in the 1990s had fielded statewide general education assessment programs abandoned them. Several states, including Missouri and West Virginia, have recently experimented with the Collegiate Learning Assessment (CLA) as the "new kid on the block," but none has as yet repeated the effort[3] —probably because these programs are expensive and maintaining them requires spending a good deal of political capital in the face of substantial institutional opposition. Results-based performance does seem to have traction when it is tied to money, a natural linkage when funding is scarce. As a consequence, results-based performance funding schemes are enjoying something of a renaissance, albeit mostly in the realm of graduation-rate incentives. Tennessee remains the lone exception in employing several learning outcomes measures in its long-established performance funding scheme.

The "public agenda" imperative became popular largely through successive issues of *Measuring Up*, the 50-state "report card" on higher education performance issued biennially since 2000 by the National Center for Public Policy in Higher Education. The basic change of perspective that *Measuring Up* embodied and promoted was from a view of higher education centered on institutions and what they do to one centered on the citizens of the state and how they benefit (Ewell & Jones, 2006). With respect to learning outcomes, this meant focusing attention less on how individual institutions perform with respect to learning and more on determining particular strengths and deficiencies in the "educational capital" of the state as a whole. *Measuring Up* consequently added a sixth "Student Learning" category to its five established grades in more measurable areas of higher education performance—and repeatedly assigned all states a grade of "Incomplete." This led to several attempts to operationalize the grade, the most prominent of which was a five-state demonstration project assessing both public and independent institutions on a range of general education examinations, licensure and admissions test results, and literacy measures (Miller & Ewell, 2005). Many states have subsequently adopted the "public agenda" model including its implied "educational capital" perspective in creating statewide goals. Although few have as yet measured anything in this vein, growing state interest in things like the Organization

[3] Some state systems of public institutions have also experimented with CLA including the California State University, the University of North Carolina, and the University of Texas systems.

for Economic Cooperation and Development's pilot test of an international student assessment in higher education suggests continuing interest. Statewide benchmark assessments in higher education comparable to K–12's National Assessment of Educational Progress (NAEP) might eventually be a legitimate outgrowth of this growing interest.

Other areas in which states retain a legitimate interest in institutional performance are less about student learning outcomes and include degree completion and job placement rates as well as evidence of effective performance in student transitions from high school to college and from one institution to another. Tracking both of these has become a good deal easier over the past two decades because of vast improvements in state student unit record database capacities (Ewell & Boeke, 2007). Measures of employment outcomes have also climbed steadily on the state policy agenda as most states have adopted a consciously articulated "grow the workforce" case for spending their scarce higher education dollars. Beyond this relatively short list, it remains in my view inappropriate for states to mandate the "how" of assessment in general education or in individual academic programs.

The Federal Government. This sector officially has only a limited role in higher education accountability. The U.S. Constitution contains no mention of federal responsibility for postsecondary education, so the major role in higher education policy making the federal government has assumed is both recent and indirect. When the Higher Education Act (HEA) of 1965 created a large and lucrative apparatus for student financial assistance, it created the need for federal oversight. Taxpayer obligations, as well as sheer prudence, demanded that federal authorities determine that the institutions administering these funds were organizationally capable of doing so. As loan monies joined direct financial assistance, the need arose to determine if the programs students bought with these funds had a decent chance of placing graduates in jobs enabling them to repay what they had borrowed. Both of these functions could be checked directly by federal reporting through occasional audits and, when funds for vocational education increased substantially in the 1970s, through job placement statistics. The need for a more general assurance of quality initiated the sometimes active, and always awkward, relationship between the federal government and accreditation.

At least two episodes of active federal involvement in higher education quality assurance that predate the Spellings Commission should be noted (Ewell, 2002). The first was the National Education Goals process that began in 1989, almost yielding a higher education equivalent of the National Assessment of Educational Progress (NAEP). The second was the battle around the 1992 HEA reauthorization, which resulted in a strong resurgence of federal authority surrounding the so-called State Postsecondary Review Entities (SPREs).[4] It is the Spellings Commission, however, which despite being short-lived and having few of its recommendations enacted, appears to signal a decisive and probably long-term shift of responsibility for quality assurance toward the federal government.

One simple reason for the growing federal role in higher education quality assurance is money. State support for public institutions (and for refilling student aid coffers in states that run such programs) is eroding. Federal dollars, meanwhile, are increasingly flowing to both public and private colleges and universities through larger federal financial aid programs (with bipartisan support and Spellings Commission recommendation) and, more recently, through the economic stimulus package. It is important to underscore the rationales behind these investments and what will probably be asked in return. The Bush administration's rationale for performance reporting was largely

The Spellings Commission, despite being short-lived and having few of its recommendations enacted, appears to signal a decisive and probably long-term shift of responsibility for quality assurance toward the federal government.

[4] Full descriptions of the history of these ill-starred efforts can be found in Ewell, 2008a, chapter 4.

couched in consumer-economy terms: potential students and their parents would use outcomes information to help them shop for a college or university, and the effects of their choices on market forces would steer institutional behavior. While sharing this concern, looming larger for the Obama administration is the role of young-adult postsecondary attainment in supporting and raising the country's competitive position internationally. In this respect, the federal government has fully embraced the "public agenda" rhetoric the states pioneered in the last decade.

What does this mean for assessment? In the near term, Congressional prohibitions will remain in effect against prescribing national standards and associated assessments erected under the leadership of Senator Lamar Alexander in the wake of the Spellings controversy and written into the Higher Education Opportunity Act (HEOA). With respect to student learning outcomes for the moment, this probably means a relatively free hand for colleges and universities so long as they clearly demonstrate responsiveness by visibly practicing assessment and publicly reporting its results. The primary exercise of the federal interest in quality assurance for the future, however, will increasingly be practiced indirectly through accreditation.

Accreditation. Regional accreditation in the U.S. began more than a century ago in an attempt to establish the boundaries of "higher" education in a rapidly expanding sector. Its programmatic counterpart developed in parallel with the rise of the professions as independent, self-regulating entities in the late 19th and early 20th centuries.[5] Among institutional accreditors, there was for a long time no specific interest in "student learning outcomes" beyond that implied by curricular structure and faculty qualifications. Among the latter, however, programmatic quality was frequently marked by more prescribed and outcomes-driven curricula and by graduate performance on established licensure examinations.

By the late 1980s, as the regional accreditors began to get interested in things like "institutional effectiveness," they gradually became a greater stimulus for institutions to practice assessment, but their treatment of the topic has historically centered much more visibly on continuous improvement than on accountability. Their requirements regarding assessment are, thus, largely about process: to ensure that the institution *has* valid and vital mechanisms for establishing and determining student achievement of key learning outcomes and to determine the extent to which the institution is *using* the resulting information to improve curricula and pedagogy.

In recent years, the role of institutional accreditation in promoting assessment has become far more complicated. On the surface—and overwhelmingly sincerely, I believe—both the rhetoric and the evolving practice of institutional accreditation is even more dedicated now than in the past to helping institutions enhance teaching and learning. Indeed, the decade preceding the Spellings Commission saw one of the most interesting and sustained periods of reform in accreditation, with not only a new focus on student learning outcomes but also with an unusually proactive effort to engage institutions through more flexible, inquiry-based review processes. Spearheaded by the Higher Learning Commission's (HLC) Academic Quality Improvement Program (AQIP), the three-stage review approach adopted by the Western Association of Schools and Colleges' (WASC) Senior Commission and the Southern Association of Schools and Colleges' (SACS) Quality Enhancement Program (QEP), these new review processes allowed institutions unprecedented latitude to focus their internal reviews on real academic challenges around which they could mobilize information and planning resources. By and large, the response was favorable and by about 2005 had considerably altered the typical compliance

In recent years, the role of institutional accreditation in promoting assessment has become far more complicated. On the surface—and overwhelmingly sincerely, I believe—both the rhetoric and the evolving practice of institutional accreditation is even more dedicated now than in the past to helping institutions enhance teaching and learning.

[5] See Ewell, 2008a, chapter 3. Note that the term "institutional accreditation" is used here to refer to accrediting organizations established to review institutions as a whole and recognized by the federal government. The most visible of these are the regional associations, but national accrediting organizations recognizing proprietary and specialized institutions have followed a similar path.

mentality that had up to that point dominated most institutions' approach to accreditation. The situation in the wake of the last administration's attempt to remake institutional accreditation as an aggressive federal quality assurance tool, however, has threatened much of this gained ground.

On the one hand, these developments are inducing institutional accreditors to take more seriously the establishment of clear standards of student academic achievement (Ewell, 2005). On the plus side, this positions them to reinforce other post-Spellings initiatives designed to re-assume the academy's responsibility for publicly assuring academic quality, such as the Voluntary System of Accountability (VSA) and the New Leadership Alliance for Student Learning and Accountability (AAC&U & CHEA, 2008). Institutional accrediting organizations remain membership associations, however, so they cannot stray too far toward establishing common standards and applying them through aggressive review. They also remain extremely limited in their ability to influence the majority of institutions not at risk of losing accreditation. The future effectiveness of institutional accreditation in both promoting good practice and in reinforcing the academy's assumption of consistent and transparent standards of student academic achievement lies entirely in the hands of the academy and its leadership.

The future effectiveness of institutional accreditation in both promoting good practice and in reinforcing the academy's assumption of consistent and transparent standards of student academic achievement lies entirely in the hands of the academy and its leadership.

"Customers," Public Opinion, and the "Fourth Estate." Much of the rhetoric of accountability in the last decade centered on "consumer information." The logic of this rhetoric is familiar: provided with concrete information about institutional conditions and performance, potential students and their parents will make individual market choices about which institutions to attend, and the fiscal (and prestige) effects of those choices will induce colleges and universities to change what they do. This particular accountability logic first appeared around the publication of graduation rates in 1989 when Congress passed the Student Right to Know and Campus Security Act (SRK), which required mandatory public reporting of these statistics. Since then, the same logic of disclosure has been written into numerous additional federal and state reporting requirements and was a major element of the Spellings Commission's assault on accreditation practices because they were alleged to be incapable of providing such information.

Despite its recent prominence, this logic's major difficulty is that it does not reflect reality. The vast majority of students in the U.S. exercise little choice about where they attend because the primary drivers of this decision are institution location and price. Even where competitive markets do exist among selective institutions in traditional higher education, information about academic performance is rarely decisive (Zemsky, 2005). The sector where the logic of this "theory of change" does seem to apply in recent years is vocationally oriented proprietary institutions, whose accreditors require them to prominently disclose standardized statistics on program completion and on related employment for graduates. These institutions would probably disclose this information anyway because they have recognized the power of performance in attracting customers. Equally important, they recognize more generally the power of marketing centered not on institutional "quality" in the traditional sense but on how students are treated in service responsiveness and in instruction tailored to individual needs. While this development does not yet pose a real threat to traditional higher education—with its elite market centered on prestige and its mass market still sufficiently subsidized that it can outcompete the for-profit sector on price—it may do so soon.

The other principal change as an accountability driver in this area is the new and growing prominence of third-party organizations in providing information about performance. The *U.S. News & World Report (USN&WR)* rankings of "America's Best Colleges" are firmly entrenched as public "markers"

of college quality and have been joined by many similar rankings publications both here and abroad. Other third-party organizations—ranging from *Measuring Up*, College Results Online (a service established by The Education Trust that reports graduation rates at four-year institutions and allows sophisticated performance comparisons among them), and a range of new performance reporting tools recently developed by Education Sector—are playing a similar role in providing performance information. Many of these nongovernmental initiatives were established to support the consumer information function central to the flawed market-based federal "theory of change." Despite this shortcoming in their original purpose, these initiatives have been shown to strongly leverage institutional behavior because they are related to the "prestige market" within which at least the top 20 percent of the institutions ranked by *USN&WR* operate. Even among nonselective institutions like community colleges, unfavorable performance on such measures as the Community College Survey of Student Engagement (CCSSE) can cause discomforting local media stories that while unlikely to influence student choice may well influence local board actions and the outcomes of important mill-levy elections.

The *USN&WR* rankings, despite unceasing criticism over more than a quarter century, have proven remarkably resistant to attack. Third-party information providers in a free market for information will always say what they want so long as it sells, suggesting an intriguing potential niche: ratings of the raters on such dimensions as the adequacy of the rater's model of quality and the soundness of the rater's metrics. This kind of analytical service has already begun emerging in the consumer products ratings industry. Another implication of the rise of the "fourth estate" as an accountability actor is the need to ensure "truth in advertising" with respect to what institutions report when they talk about performance—one reason why institutional accreditors have increasingly adopted "integrity" standards governing how institutions portray themselves through catalogues, recruitment materials, and web sites.

Managing the Tension: Some Principles of Response

In this section, I offer four principles to guide institutions' responses to the increasingly strident calls for external accountability while preserving and developing the institutions' internal capacity for evidence-based continuous improvement.

1. Respond visibly to domains of legitimate external concern. The late 1980s was a period when many states were spending relatively heavily on higher education in the name of developing human capital and workforce skills. Providing stakeholders with reliable information about things like graduation and job placement rates made good sense then, just as it does today. But the current external demand is more about providing direct evidence of student academic achievement—centered on broad undergraduate skills like critical thinking, communication, and problem solving than about where graduates go. And the major client is different: the federal government, acting through recognized accrediting organizations, purportedly on behalf of "customers."

Much also turns on what is meant by "legitimate." Twenty years ago, the very notion that government actors could demand information about performance from self-governed—and, especially, independent—institutions was considered illegitimate by many members of the academy. Indeed, the primary recommendation I made to institutions at that time was that they respond with something other than resistance. As noted earlier, our community has moved on from this position and virtually all colleges and universities are currently engaged in some kind of assessment. The question now is exactly what types of learning outcomes information can be considered "legitimate." One product of the Spellings adventure was that responsibility for setting learning objectives rests with institutions themselves, not with external actors.

Four guiding principles for responding to calls for external accountability that also preserve and develop institutional capacity for evidence-based continuous improvement:

1. *Respond visibly to domains of legitimate external concern.*
2. *Show action on the results of assessment.*
3. *Emphasize assessment at the major transition points in a college career.*
4. *Embed assessment in the regular curriculum.*

How institutions choose to discharge this responsibility remains up to them. Many colleges and universities hold the position that such information should be comparable across institutions, a position reflected in the VSA; but many do not do so, maintaining that each institution should be free to set its own learning goals and assessment methods. A parallel discord surrounds the term "respond." The VSA is unambiguous in construing "response" to be complete public disclosure, while many institutions continue to construe "response" to mean conducting a program of assessment that measures up to accreditation standards, regardless of whether the public ever gets to see the results.

Proactive response has an additional implication: the *ways* and the *tone* in which outcomes information are reported are at least as important as the content of the disclosures themselves. The rhetorical standpoint from which higher education responds to demands for outcomes-based accountability is thus extremely important. When acquiescence has occurred in the past, it was largely with an attitude of compliance. Even when accountability requirements were seen by college leaders as necessary—part of the price that colleges and universities had to pay to receive public funds—their acceptance was largely grudging and unenthusiastic. In contrast, higher education's posture of engagement about assessment's accountability dimension today must emphasize a commitment to a *collective responsibility* for teaching and learning and their results. Instead of seeing assessment as an aspect of higher education's responsibility to its *funders*—legitimate though this may be—both faculty and academic leaders need to see it as part of our accountability to *ourselves*. This is, after all, how we operate in the realm of research, and it is why mechanisms like independent peer review are so important to maintaining scholarly integrity. It needs to happen in teaching and learning as well.

Far too many institutions, dominated by the need to respond to external actors like states or accreditors, approach the task of assessment as an act of compliance, with the objective being simply to measure something and the exercise ending as soon as the data are reported.

Some evidence that the academy has already taken this advice is provided by the recent launch of the New Leadership Alliance for Student Learning and Accountability (the "Alliance"). This organization grew out of informal gatherings of higher education leaders in the midst of the HEA reauthorization debates some two years ago as an effort to anticipate calls for public accountability and to develop a proactive response. Advancing principles that call for institutions to set rigorous goals for undergraduate learning, to collect evidence of whether these goals are being achieved, and to act on this information continuously to improve instruction (see AAC&U & CHEA, 2008), the Alliance was formally launched six months ago. The Alliance plans several initiatives, the foremost of which is formal certification of institutions that are doing an exceptional job of acting on the principles of assessment and continuous improvement as determined by an independent, credible, third-party review process. But the real work is up to the institutions. If enough institutions go along with it, especially those in the high-prestige sector that do well in *USN&WR* rankings, it might actually work this time.

2. Show action on the results of assessment. When I offered this advice in 1987, I sincerely believed that sincere and consistent action to assess and improve on the part of a significant number of colleges and universities would be sufficient to "solve" the academy's accountability problem. I argued then that "much of the recent concern about higher education's performance is due more to a loss of confidence than a desire on the part of external authorities to actively run the enterprise" (Ewell, 1987, p. 14). If the academy were only to show sincere and concerted action on the assessment-for-improvement agenda, accountability pressures might let up.

Now I am not so sure. The new policy centrality of higher education and the nation's slipping competitive position internationally means that today's accountability problem goes beyond just lack of responsiveness to a crisis in specific performance. Accordingly, colleges and universities will not only have to demonstrate sincere efforts to improve student learning but will also have to prove that their students are achieving at adequate levels in the first place.

This will increasingly mean reporting actual learning outcomes in comparative or benchmarked forms as well as being transparent about internal efforts at continuous improvement—actions the Alliance champions.

Experience has shown that implementing continuous improvement is not all that easy either, however. A major reason why higher education's accountability problem has not gone away is certainly that most institutions lack the will to "show action" by closing the loop on assessment results. Many institutions simply do not know, however, how to implement evidence-based continuous improvement. Why is this? One reason is the general nature of most assessment results when compared to the concrete realities of changing curriculum and pedagogy. Although exceptions are apparent, most assessment evidence is simply not fine grained enough to yield actionable information at this level. A similar reason is that such evidence tends to be presented in the form of central tendency measures, which don't show the patterns of strength and weakness or the variations in performance across types of students needed to guide intervention (Kuh, 2007). Finally, information about outcomes alone doesn't tell faculty what to fix. In contrast, tools that focus on institutional climate and staff behaviors, like NSSE, are far more amenable to institutional action.

For assessment-based improvement to work effectively learning objectives must be inescapable: They are in catalogues, on syllabi, and visible in the criteria faculty use to assign grades.

Despite these challenges, experience suggests a number of tricks of the trade that savvy institutions have learned about how to use assessment evidence for improvement. One of the most important of these is thinking about utilization from the beginning, as part of the assessment design itself. Far too many institutions, dominated by the need to respond to external actors like states or accreditors, approach the task of assessment as an act of compliance, with the objective being simply to measure something and the exercise ending as soon as the data are reported. Best practice, on the other hand, suggests that involving faculty formally, and in detail, in exercises designed to craft a set of specific teaching-related questions that faculty want answered can yield substantial dividends. Such questions are frequently best framed in terms of particular pedagogical challenges that real faculty face in real classrooms. Up-front planning for utilization should also include expectations exercises: What do participants expect the data to reveal, or what might be the action consequences of this or that result? Another important technique here is to successively disaggregate results for specific populations or outcomes dimensions. As noted earlier, averages tell a user very little about what is actually going on and, therefore, what needs to be changed.

Especially important for continuous improvement, though, is the need to create concrete opportunities for faculty and staff to look at such disaggregated data *together* to discuss what the data mean and to consider what particular action implications they might have. Far too few opportunities for such thoughtful collective reflection about evidence are available on today's college campuses, and institutional leaders need to go out of their way to create them. The best examples of effective utilization of assessment results begin with sharing a few carefully chosen pieces of information with a faculty-staff group charged with examining and discussing them. This process usually results in a few initial conclusions, but far more additional questions that need to be answered by further analysis of the available data. This iterative process is far more helpful in informing action than a "data dump" containing few clues as to what is important or practicable. The analytical art of designing such conversations to shape a productive discussion is in short supply now, just as it was two decades ago.

For assessment-based improvement to work effectively, moreover, learning objectives must be inescapable: They are in catalogues, on syllabi, and visible in the criteria faculty use to assign grades. Neither faculty nor students can forget they are there and the institution's commitment to standards and continuous improvement should be apparent in virtually every action the institution takes with respect to teaching and learning. The same holds true for important

administrative processes like mission review and strategic planning, program review and curricular development, budgeting, and governance. Institutions demonstrate their commitment to evidence-based management and improvement to their members and their stakeholders by constantly referencing—and taking seriously—their goals for learning and the evidence they have about the extent to which these goals are being achieved. The rhetoric of institutional leaders has a lot to do with this. To credibly act on accountability, as well as to reinforce local improvement, presidents and provosts need to explicitly reference assessment and its results whenever they announce important decisions or communicate with the public. And, as noted earlier, this stance must not be grudging but proactive.

3. Emphasize assessment at the major transition points in a college career. If an institution's goals for student learning are truly dominant, they must permeate the entire curriculum and be explicitly assessed at multiple points in a student's career. This means being sure of college readiness on entry and requiring clear demonstrations of mastery to earn a degree. It also requires attention to what happens between these two points, which is a good deal more difficult to define. AAC&U's Liberal Education and America's Promise (LEAP) initiative is an example of exactly what is required in its call for institutions to create "milestone" assessments as students progress in their programs as well as "capstone or culminating experiences" emphasizing the integration of their knowledge in their majors (AAC&U, 2008, p. 8).

One early transition is a good place to start: post-testing basic skills as an exit requirement for developmental education. The logic behind this advice is that this is less threatening to mainstream faculty than directly looking at their own effectiveness in college-level teaching. After all, if it is legitimate to place students appropriately using testing tools like COMPASS or AcuPlacer, it should be equally appropriate to use the same tools to determine if they mastered these skills after remediation. Like the results of any assessment, the data such post-testing produced could be analyzed and fed back to developmental instructors to inform improvement, thus "rehearsing" the assessment-improvement process in a way less risky than immediately turning assessment technology loose on general education.

Equally important, developmental education is itself a major national challenge. Students are being placed into developmental study in record numbers, and this bottleneck in the "education pipeline" is one of the major reasons why U.S. competitive performance with respect to young-adult postsecondary degree attainment is slipping. Assessment constitutes an essential part of the solution to this challenge and, in some places, is being used appropriately and effectively. Fourteen states now employ common placement tests to govern entry into college-level work, 12 of which have also set common cut scores (Ewell, Boeke, & Zis, 2008). This begins to address the national need to establish learning-based benchmarks of college readiness—a badly needed step. Far less progress has been made in using the results of placement testing—or other assessment, for that matter—to improve the effectiveness of developmental education itself. New approaches to instructional delivery in this arena are springing up constantly including modular short-course designs with content tailored individually to address particular assessed deficiency, "just-in-time" provision, blended technology-based and face-to-face formats, and contextualized courses in which basic skills content is embedded in particular areas of study. These all beg to be evaluated and improved using systematic data on posttest-based outcomes.

The second obvious "transition point" at which to locate assessment—and the one most germane to accountability purposes—is at the conclusion of a program. External stakeholders above all want evidence that those who hold

External stakeholders above all want evidence that those who hold a degree have actually attained a clearly defined level of performance with respect to knowledge and skills.

a degree have actually attained a clearly defined level of performance with respect to knowledge and skills. Indeed, for some of these stakeholders—employers, for example—absolute attainment may be all they want to know. The best illustration here is the many licensure or certification examinations that regulate entry into professional fields ranging from teacher education to the health professions—often used as benchmarks of quality for academic programs. Another example is provided by the few state or system-specific programs that use "gate-keeping" testing, such as the South Dakota Board of Regents use of ACT's CAAP or the California State University's Writing Proficiency Examination that is required of all students to graduate. For those funding (or potentially enrolling in) an institution or program, though, the further question of "value-added" may arise: How much gain in designated knowledge or skill has a student experienced as a result of attendance? Institutional assessment programs tilted heavily toward accountability that put substantial weight on national instruments like CAAP or CLA, for example, attempt to capture both the performance of seniors at or near their point of graduation and that of students just starting out. Disaggregating the results of such summative assessments and looking at patterns of variation as well as mean scores, as noted earlier, may simultaneously aid institutional improvement efforts by identifying gaps in performance among particular population groups or by suggesting further things to investigate.

More suited to a balance between accountability and improvement purposes is a culminating demonstration—usually in the major field—in the form of a comprehensive examination, external examiner, senior thesis or project, or capstone course.

More suited to a balance between accountability and improvement purposes is a culminating demonstration—usually in the major field—in the form of a comprehensive examination, external examiner, senior thesis or project, or capstone course. In the traditional liberal arts curricula of 50 years ago, such features were common—especially in selective liberal arts colleges; but most had disappeared by the 1980s. As assessment became increasingly salient, many colleges and universities took a second look at them. Capstone courses, in particular, were appealing not only for their instructional value as an opportunity to integrate prior knowledge but also as convenient settings for assessment, and they have become quite common as a result. Not every course labeled "capstone" measures up to what is needed for assessment, however, unless it is intended solely as a convenient opportunity to administer an unrelated examination like an ETS Major Field Achievement Test. To be truly suitable, in fact, a capstone should meet three conditions. First, key assignments in the course ought to examine knowledge and abilities taught throughout the program. This means, under most circumstances, that a specialized course in the discipline would not be suitable. Second, ratings of student performance ought to be the product of multiple expert readers or faculty. This means, under most circumstances, that all the grading in the course should not be done by one individual. Finally, the assignments or demonstrations used for assessment in the course—and the rating schemes used to look at student responses—should be capable of generating information on multiple dimensions of student performance faculty have deemed important. Although these may be difficult conditions to meet fully, many institutions have found ways to approximate them, and it remains important to state them in their purest form as an "ideal type" to guide design.

If the beginning and the end of a program are obvious places to assess, the "middle" in the AAC&U schema has received a good deal less attention. To be sure, some institutions and states have historically used the end of the sophomore year as an occasion to directly assess student competency in general education skills. The most prominent example was the College Level Academic Skills Test (CLAST), which was used as a "rising junior" examination in Florida for many years. A more creative example is the King's College Sophomore/Junior Diagnostic Project, in which each student demonstrates a specific set of general education skills in the context of the major that he or she is about to enter (Farmer, 1988). Another logical place to examine midpoint skills is at the point of transfer between two-year and four-year institutions. For example, several states including Missouri and Utah have tried to cast

their transfer requirements on the basis of "abilities" in addition to just course requirements, although none has as yet created assessments as an alternative way to meet transfer requirements (Ewell, Boeke & Zis, 2008). By and large, it has proven logistically difficult for most institutions to implement "middle-level" assessments as an "add-on" to the regular curriculum, and this is why the final, following, principle is so important.

4. Embed assessment in the regular curriculum. In the early days of the assessment movement, campus assessment practices were consciously separated from what went on in the classroom. This separation helped increase the credibility of the generated evidence because, as "objective" data-gathering approaches, these assessments were free from contamination by the subject they were examining. Partly as a result, assessment practitioner rhetoric at the time strongly criticized grades as a valid and reliable measure of student learning. Instead, most early-adopters relied on off-the-shelf commercial tests like the ACT Assessment or the GRE, originally designed for something else. My original call to embed assessment in the regular curriculum was intended to shift the dominant pattern of thinking away from this "exo-skeletal" architecture toward a reintegration of assessment with the everyday practices of teaching and learning, and I remain committed to this today. Far too many institutions have established learning outcomes in response to accreditation requirements and to drive assessments without ensuring that these goals are continuously mapped to, and reinforced by, the teaching and learning process *throughout* the curriculum as part of a systematic competency-based approach.

Far too many institutions have established learning outcomes in response to accreditation requirements and to drive assessments without ensuring that these goals are continuously mapped to, and reinforced by, the teaching and learning process throughout the curriculum as part of a systematic competency-based approach.

One demonstrated approach to embedding assessment in the regular curriculum is the "competency growth plan" for general education assessment pioneered by King's College (Farmer, 1988), which explicitly uses specific assignments in each academic program to certify achievement at successively more challenging levels up to the senior year. These assignments thus move beyond their course-specific application designed by a single faculty member and become the "collective property" of the faculty as a whole. Because of their collective importance, moreover, student responses to these assignments are rated independently of the regular grading process, using specially design rubrics or scoring guides. Obviously, the number of assignments designated for this purpose must be kept under control, so most institutions that adopt this approach have established no more than one or two midpoint levels of achievement on the way to a degree.

As assessment evidence, student work samples produced in response to regular course assignments have the virtue of having been already generated and are considered considerably more "authentic" by most faculty members than externally produced measures. Probably more important, the problem of student motivation to do well on an exercise that does not count is essentially off the table. As many institutions have found, however, harnessing student work as assessment evidence is harder than it looks. Assignments need to be carefully designed to elicit responses appropriate for consistent scoring, scoring rubrics need to be developed that yield reliable ratings across graders, and a mechanism needs to be in place to assemble and store the artifacts themselves.

When King's College and Olivet College introduced this embedded approach 20 years ago, the logistical tasks entailed were nearly overwhelming. These tasks are much simpler now due to recent advances in managing assessment results through technology (Hutchings, 2009). Some of these technology tools, like TracDat, provide easy-to-access databases of assessment results drawn from multiple sources that can be used to generate summary reports as well as more detailed breakdowns of student performance. Some, like E-Lumen, provide a platform that allows faculty to collectively build and test rubrics associated with particular assignments and to score student work collectively online. The overwhelming method of choice in this arena, however, has become the electronic portfolio. National organizations such as AAC&U have promoted the

widespread use of electronic portfolios particularly because they promise to simultaneously support the detailed inquiry into student learning that assessment for improvement requires while retaining the capacity to report a few summative measures.

Embedded assessment approaches like competency-anchored assignments and electronic portfolios are becoming increasingly comfortable for faculty and can now be implemented efficiently on a large scale. They represent what is today the most compelling route that an institution can take to navigate the accountability-improvement tension.

Concluding Thoughts

Despite many changes in motive and circumstances, the tension between the purposes of assessment for accountability and assessment for improvement that characterized the higher education landscape 20 years ago continues to exist today. Giving too much attention to accountability risks losing faculty engagement—effectively suppressing the sustained, critical self-examination that continuous improvement demands. Devoting attention solely on the internal conversations needed for improvement, on the other hand, invites external actors to invent accountability measures that are inappropriate, unhelpful, or misleading. Managing this tension requires staking out a middle ground. One of the promises of using institutional accreditation as the primary vehicle for stimulating assessment and discharging accountability is that it tries to do just this. Carefully distinguishing processes and initiatives designed to accomplish one or the other of these purposes so they do not interfere with one another can often be useful. Several regional accreditors, for example, have designed their approaches to consciously separate "compliance" and "deep engagement" activities so that the former can be accomplished without driving out the latter. Inside campus practice, the embedded assignment and portfolio approaches to assessment also represent a middle ground because they can generate and capture the necessary informational detail to inform curricular improvement while, if designed properly, retaining the capacity to produce summary benchmarks of student academic achievement for departments or the institution as a whole.

Although the tension remains, some things today are decidedly different. As noted already, I was convinced 20 years ago that widespread institutional attention to designing robust assessment-for-improvement programs and to taking visible action based on evidence from these programs would be sufficient to provide accountability—and would obviate the need to report measures of student achievement that can be benchmarked or compared. I do not believe this today. We will need to do both in the coming years, buying the necessary time to accomplish the more needed and desirable task of building institutional infrastructures for evidence-based continuous improvement with public performance reporting (Ewell, 2008b). Another difference compared to 20 years ago is this conversation's urgency today, which means that attention to it has become unavoidable. Some elements of the accountability-improvement tension will no doubt prove sufficiently long-lived that our successors will be revisiting them two decades hence, but we can all hope that the present urgency results this time in our getting much more done.

Some elements of the accountability-improvement tension will no doubt prove sufficiently long-lived that our successors will be revisiting them two decades hence, but we can all hope that the present urgency results this time in our getting much more done.

References

Astin, A., Banta, T., Cross, P., El-Khawas, E., Ewell, P., Hutchings, P., Marchese, T., McClenney, K., Mentkowski, M., Miller, M., Moran, E., & Wright, B. (1992). *Principles of good practice for assessing student learning*. AAHE Bulletin, 45, 4.

Association of American Colleges and Universities (AAC&U). (2008). *Our students' best work: A framework for accountability worthy of our mission* (2nd ed.). Washington, DC: Author.

Association of American Colleges and Universities (AAC&U) & Council for Higher Education Accreditation (CHEA). (2008). *New leadership for student learning and accountability: A statement of principles, commitment to action*. Washington, DC: Authors.

Chickering, A. W., & Gamson, Z. S. (1987, March). Seven principles for good practice in undergraduate education. *AAHE Bulletin*, 39(7), 3–7.

Ewell, P. T. (1987). *Assessment, accountability and improvement: Managing the contradiction*. Washington, DC: American Association for Higher Education.

Ewell, P. T. (2002). An emerging scholarship: A brief history of assessment. In T. W. Banta & Associates (Eds.), *Building a scholarship of assessment* (pp. 3–25). San Francisco, CA: Jossey-Bass.

Ewell, P. T. (2005). Can assessment serve accountability? It depends on the question. In J. C. Burke & Associates (Eds.), *Achieving accountability in higher education* (pp. 1–24). San Francisco, CA: Jossey-Bass.

Ewell, P. T. (2008a). *U.S. accreditation and the future of quality assurance: A CHEA tenth anniversary report*. Washington, DC: Council for Higher Education Accreditation.

Ewell, P. T. (2008b). Assessment and accountability in America today: Background and context. In V. M. H. Borden & G. Pike (Eds), *Assessing and accounting for student learning: Beyond the Spellings Commission* (New Directions for Institutional Research, Assessment Supplement 2007, pp. 7–18). San Francisco, CA: Jossey-Bass.

Ewell, P. T., & Boeke, M. (2007). *Critical connections: Linking states' unit record systems to track student progress*. Indianapolis, IN: Lumina Foundation for Education.

Ewell, P. T., Boeke, M., & Zis, S. (2008). *State policies on student transitions: Results of a fifty-state inventory*. Boulder, CO: National Center for Higher Education Management Systems.

Ewell, P. T., & Jones, D. P. (2006). State-level accountability for higher education: On the edge of a transformation. In N. B. Shulock (Ed.), *Practitioners on making accountability work for the public* (New Directions for Higher Education No. 135, pp. 9–16). San Francisco, CA: Jossey-Bass.

Farmer, D. W. (1988). *Enhancing student learning: Emphasizing essential competencies in academic programs*. Wilkes-Barre, PA: King's College.

Hutchings, P. (2009, May-June). The new guys in assessment town. *Change*, 41(3), 26–33.

Kuh, G. D. (2007). Risky business: Promises and pitfalls of institutional transparency. *Change*, 39(5), 30–35.

Miller, M. A., & Ewell, P. T. (2005). *Measuring up on college-level learning*. San Jose, CA: National Center for Public Policy and Higher Education.

National Center for Public Policy in Higher Education (NCPPHE). (2000–). *Measuring up: The fifty-state report card on higher education*. San Jose, CA: Author.

Organization for Economic Cooperation and Development (OECD). (2008). *Education at a glance: OECD indicators*. Paris, France: Author.

U. S. Department of Education. (2006). *A test of leadership: Charting the future of American higher education* (Report of the commission appointed by Secretary of Education Margaret Spellings). Washington, DC: Author.

Zemsky, R. M. (2005). The dog that doesn't bark: Why markets neither limit prices nor promote educational quality. In J. C. Burke & Associates (Eds.), *Achieving accountability in higher education* (pp. 275–295). New York, NY: Jossey-Bass.

NILOA National Advisory Panel

Trudy W. Banta
Professor
Indiana University-Purdue University Indianapolis

Douglas C. Bennett
President
Earlham College

Robert M. Berdahl
President
Association of American Universities

Molly Corbett Broad
President
American Council on Education

Judith Eaton
President
Council for Higher Education Accreditation

Richard Ekman
President
Council of Independent Colleges

Joni Finney
Practice Professor
University of Pennsylvania

Vice President, National Center for Public Policy and Higher Education

Susan Johnston
Executive Vice President
Association of Governing Boards

Paul Lingenfelter
President
State Higher Education Executive Officers

George Mehaffy
Vice President
Academic Leadership and Change
American Association of State Colleges and Universities

Margaret Miller
Professor
University of Virginia

Charlene Nunley
Program Director
Doctoral Program in Community College Policy and Administration
University of Maryland University College

Randy Swing
Executive Director
Association for Institutional Research

Carol Geary Schneider
President
Association of American Colleges and Universities

David Shulenburger
Vice President
Association of Public and Land-Grant Universities

Belle Wheelan
President
Southern Association of Colleges and Schools

George Wright
President
Prairie View A&M University

Ex-Officio Members

Peter Ewell
Vice President
National Center for Higher Education Management Systems

Stanley Ikenberry
Interim President
University of Illinois

George Kuh
Chancellor's Professor
Indiana University

National Institute for Learning Outcomes Assessment (NILOA) Mission

NILOA's primary objective is to discover and disseminate ways that academic programs and institutions can productively use assessment data internally to inform and strengthen undergraduate education, and externally to communicate with policy makers, families and other stakeholders.

NILOA Occasional Paper Series

NILOA Occasional Papers are commissioned to examine contemporary issues that will inform the academic community of the current state-of-the art of assessing learning outcomes in American higher education. The authors are asked to write for a general audience in order to provide comprehensive, accurate information about how institutions and other organizations can become more proficient at assessing and reporting student learning outcomes for the purposes of improving student learning and responsibly fulfilling expectations for transparency and accountability to policy makers and other external audiences.

Comments and questions about this paper should be sent to sprovez2@illinois.edu.

About NILOA

- The National Institute for Learning Outcomes Assessment (NILOA) was established in December 2008.
- NILOA is co-located at the University of Illinois and Indiana University.
- The NILOA web site went live on February 11, 2009.
  www.learningoutcomesassessment.org
- The NILOA research team reviewed 725 institution web sites for learning outcomes assessment transparency from March 2009 to August 2009.
- One of the co-principal NILOA investigators, George Kuh, founded the National Survey for Student Engagement (NSSE).
- The other co-principal investigator for NILOA, Stanley Ikenberry, was president of the University of Illinois from 1979 to 1995 and of the American Council of Education from 1996 to 2001. He is currently serving as Interim President of the University of Illinois.
- Peter Ewell joined NILOA as a senior scholar in November 2009.

NILOA Staff

NATIONAL INSTITUTE FOR LEARNING OUTCOMES ASSESSMENT

Stanley Ikenberry, *Co-Principal Investigator*

George Kuh, *Co-Principal Investigator and Director*

Peter Ewell, *Senior Scholar*

Staci Provezis, *Project Manager*

Jillian Kinzie, *Associate Research Scientist*

Jason Goldfarb, *Research Analyst*

Natasha Jankowski, *Research Analyst*

Gloria Jea, *Research Analyst*

Julia Makela, *Research Analyst*

NILOA Sponsors

Carnegie Corporation of New York

Lumina Foundation for Education

The Teagle Foundation

Produced by Creative Services | Public Affairs at the University of Illinois for NILOA.
10.032

National Institute for Learning Outcomes Assessment

For more information, please contact:

National Institute for Learning Outcomes Assessment (NILOA)
University of Illinois at Urbana-Champaign
340 Education Building
Champaign, IL 61820

learningoutcomesassessment.org
sprovez2@illinois.edu
Fax: 217.244.3378
Phone: 217.244.2155