An Emerging Scholarship: A Brief History of Assessment

Peter T. Ewell

National Center for Higher Education Management Systems (NCHEMS)

Draft

This chapter offers a brief historical and analytical review of the assessment movement from approximately 1985 to date. It first examines some major events and forces influencing assessment’s evolution as a “scholarship,” including demands for curricular and pedagogical reform, shifting patterns of accountability, and changes in instructional delivery. It also examines significant scholarly themes and issues that have arisen in assessment’s short history in such realms as epistemology, methodology, politics, and the use of information. The chapter concludes that assessment scholarship has become rich, robust, and strong. Whether it can or should continue as a distinct conversation outside the mainstream of higher education is more debatable.

Forerunners

The intellectual roots of assessment as a scholarship extend well before its emergence as a recognizable movement. Some of its most visible forbears are about undergraduate learning and the student experience in college. Others helped ground its conscious orientation toward action and improvement. Methods and techniques drawn from these established traditions decisively influenced the language and methods of early assessment practitioners and continue to do so today.

• Student Learning in College. This research tradition examines collegiate learning as a particular application of educational and developmental psychology. As such, its primary
objective is discipline-based hypothesis testing and theory building, though its authors have often drawn implications for practice. Some of this work dates back to the 1930s and 40s (e.g., Learned and Wood 1938), and much of it focused on single colleges enrolling 18- to 21-year-old students in traditional residential environments. General maturation and attitudinal development was thus as much an area of interest as cognitive gain (Chickering 1969). By the end of the 1960s there was a large enough body of work in this area for Feldman and Newcomb (1969) to synthesize its findings, updated some two decades later by Pascarella and Terenzini (1991). On the verge of assessment’s emergence in the late 1970s, a trio of volumes was especially influential: Astin’s *Four Critical Years* (1977) established the metaphor of “value-added” and promoted the use of longitudinal studies to examine net effects; Bowen’s *Investment in Learning* (1977) helped establish a public policy context for assessment by emphasizing the societal returns on investment associated with higher education; and Pace’s *Measuring the Outcomes of College* (1979) emphasized the role of college environments and actual student behaviors. The contributions of this research tradition to assessment were both conceptual and methodological. Among the most prominent were basic taxonomies of outcomes, models of student growth and development, and tools for research like cognitive examinations, longitudinal and cross-sectional surveys, and quasi-experimental designs.

- **Retention and Student Behavior.** Closely related to research on college student learning, a distinct literature on retention emerged in the late 1960s and 70s and had some very specific impacts on assessment practice. First, it quickly organized itself around a powerful theoretical model—Tinto’s notion of academic and social integration (Tinto 1975)—which proved equally useful in guiding applied research on student learning (e.g., Terenzini,
Pascarella and Lorang 1982). Second, the phenomenon of student attrition constituted an ideal proving ground for new methodologies involving longitudinal study designs, specially-configured surveys, and multivariate analytical techniques, later adopted by many assessment practitioners. Third and perhaps decisively, retention scholarship was action research: though theoretically grounded and methodologically sophisticated, its object was always informed intervention (e.g., Lenning, Beal and Sauer 1980). Together, these features yielded an excellent model of applied scholarship that, consciously or unconsciously, many assessment practitioners worked to emulate.

- **Evaluation and “Scientific” Management.** The 1960s and 70s also saw the rise of program evaluation as an action research tradition. Occasioned by the many large-scale federal programs launched at that time, program evaluation first relied largely on quantitative methods. It was also related to a wider movement toward “scientific” management that quickly found applications in higher education in the form of strategic planning, program review, and budgeting. The kind of “systems thinking” embedded in this tradition demanded explicit attention to student outcomes (e.g., Enthoven 1970) in order to provide a needed “output variable” for cost-benefit studies and investigations of social return on investment. This tradition also yielded one of the most extensive taxonomies of collegiate outcomes ever produced (Lenning, Lee, Micek and Service 1977) and stimulated a range of surveys designed to provide campuses with information about how students used and perceived their programs. Literature drawn from program evaluation further provided assessment with a ready-made set of models and vocabularies (e.g., Light, Singer and Willett 1990). Somewhat later program evaluation began to embrace more qualitative methods (e.g., Guba and Lincoln 1981). These more “authentic” approaches, which emphasized holistic examination of
organizational situations and often employed open-ended interviewing and participant-observation, also provided an early language for assessment for those skeptical of overly empirical methodologies

- **Mastery Learning.** The mastery and competency-based learning movement began in elementary and secondary education, but quickly found postsecondary applications in adult and professional education by the mid-1960s. Because mastery-based designs for learning are entirely based on agreed-upon outcomes, assessing and certifying individual student achievement was always paramount. A related development was the assessment of prior learning. Corporate assessment centers, meanwhile, were developing ways to examine and certify complex higher-order abilities by observing group and individual performances on authentic tasks (Thornton and Byham 1982). Collectively, these traditions provided the conceptual foundation for “alternative” institutions like Empire State, Evergreen State, Regents College, Antioch College, and the School for New Learning at DePaul, and, by far the most influential, Alverno College (Alverno College Faculty 1979). They also yielded a cadre of early assessment practitioners, skilled in evaluating student portfolios and other authentic evidence of student attainment. Two contributions were especially important for the early assessment movement. First, mastery methods posed an effective alternative to the prominent (and politically popular) “testing and measurement” paradigm. Second, they could boast a track record that proved that assessment in higher education was not just a popular “theory”; it could actually be done.

These four practice traditions and their associated literatures are quite different and only a few in the early 1980s were reading them all. More significantly, their values and methodological traditions are frequently contradictory, revealing conceptual tensions that have fueled assessment
discussions ever since. One is a clash of guiding metaphor between quantitative “scientific” investigation and qualitative “developmental” observation. Another addresses how assessment is positioned in the teaching-learning process: the “evaluation” and “measurement” traditions consciously divorce the process of investigating student attainment from the act of instruction in the name of objectivity; “mastery” traditions, in contrast, consider the two inseparable. A final distinction concerns the predominant object of assessment—whether its principal purpose is to examine overall program/institutional effectiveness or to certify what a particular student knows and can do. As any examination of early assessment citations will show, all four traditions helped shape language and practice in the early 1980s. What is surprising in retrospect is that such disparate scholarly traditions could be related at all and that they continue to inform such a lively scholarship.

**Birth of a Movement**

While no one has officially dated the birth of the “assessment movement” in higher education, it is probably safe to propose the First National Conference on Assessment in Higher Education held in Columbia, SC in the fall of 1985. Co-sponsored by the National Institute of Education (NIE) and the American Association for Higher Education (AAHE), the origins of this conference vividly illustrate the conflicting political and intellectual traditions that have been with the field ever since. The proximate stimulus for the conference was a report called *Involvement in Learning* (NIE 1984). Three main recommendations formed its centerpiece, strongly informed by research in the student learning tradition. In brief, they were that higher levels of student achievement could be promoted by establishing high expectations for students, by involving them in active learning environments, and by providing them with prompt and useful feedback. But the report also observed that colleges and universities could “learn” from
feedback on their own performances and that appropriate research tools were now available for them to do so.

This observation might have been overlooked were it not consistent with other voices. One set came from within the academy and focused on curriculum reform, especially in general education. Symbolized by other prominent reports like *Integrity in the College Curriculum* (AAC 1985) and *To Reclaim a Legacy* (Bennett 1984), their central argument was the need for a coherent curricular experiences which could best be shaped by ongoing monitoring of student learning and development. From the outset in these discussions, the assessment of learning was presented as a form of “scholarship.” Faculties ought to be willing to engage in assessment as an integral part of their everyday work. A concomitant enlightened, but unexamined, assumption was that the tools of social science and educational measurement, deployed appropriately, could be adapted by all disciplines to further this process of ongoing inquiry and improvement.

A second set of voices arose simultaneously outside the academy, consisting largely of state-based calls for greater accountability. In part, these calls were a byproduct of the far more visible attention then being paid to K-12 education, symbolized by the U.S. Department of Education’s (1983) report *A Nation at Risk*. In part, it stemmed from a renewed activism by governors and legislatures, based on their growing recognition that postsecondary education was a powerful engine for economic and workforce development. Both themes were apparent in yet another national report—revealingly titled *Time for Results* (NGA 1986). As it was being issued, states like Colorado and South Carolina adopted assessment mandates requiring public colleges and universities to examine learning outcomes and report what they found. (A few other states like Tennessee and Florida had, for varying reasons, been doing assessment for several years, using common standardized tests). By 1987 when the first stock-taking of this growing policy
trend occurred (Boyer, Ewell, Finney, and Mingle 1987) about a dozen states had similar mandates. By 1989, this number had grown to more than half (Ewell, Finney, and Lenth 1990).

Given this history, the motives of those attending the first national assessment conference were understandably mixed. Many were there under the banner of *Involvement in Learning*, seeking reasonable and valid ways to gather information to improve curriculum and pedagogy. At least as many (and probably more) were there in response to a brand new mandate. Clear to all were the facts that they had few available tools, only a spotty literature of practice, and virtually no common intellectual foundation on which to build. Filling these yawning gaps in the period 1985-88 was a first and urgent task for the scholarship of assessment. In beginning this task, practitioners faced three major challenges:

- **Definitions.** One immediate problem was that the term “assessment” meant different things to different people. Initially, at least three meanings and their associated traditions of use had therefore to be sorted out. The most established had its roots in the mastery-learning tradition, where “assessment” referred to the processes used to determine an individual’s mastery of complex abilities, generally through observed performance (e.g., Alverno College Faculty 1979). Adherents of this tradition emphasized development over time and continuous feedback on individual performance—symbolized by the etymological roots of the word “assessment” in the Latin *ad + sedere*, “to sit beside” (Loacker, Cromwell and O’Brien 1986). A far different meaning emerged from K-12 practice, where the term described large-scale testing programs like the federally-funded National Assessment of Educational Progress (NAEP) and a growing array of state-based K-12 examination programs. The primary object of such “large-scale assessment” was not to examine individual learning but rather to benchmark school and district performance in the name of
accountability. Its central tools were standardized examinations founded on well-established psychometric principles, designed to produce summary performance statistics quickly and efficiently. Yet a third tradition of use defined “assessment” as a special kind of program evaluation, whose purpose was to gather evidence to improve curricula and pedagogy. Like large-scale assessment, this tradition focused on determining aggregate not individual performance, employing a range of methods including examinations, portfolios and student work samples, surveys of student and alumni experiences, and direct observations of student and faculty behaviors. An emphasis on improvement, moreover, meant that assessment was as much about using the resulting information as about psychometric standards.

All three definitions raised explicitly the dichotomy of purpose apparent from the outset: accountability vs. improvement. Other differences addressed methods and units of analysis—essentially whether quantitative or qualitative methods would predominate and whether attention would be directed largely toward aggregate or individual performance. Clarifying such distinctions in the form of taxonomies helped sharpen initial discussions about the meaning of “assessment” (Terenzini 1989). They also helped further a terminological consensus centered on the use of multiple methods for program improvement (AAHE 1992).

- **Instruments.** A second challenge faced by early assessment practitioners was to quickly identify credible and useful ways of gathering evidence of student learning. Virtually all the available instruments were designed for something else. Ranging from admissions tests like the ACT Assessment and the Graduate Record Examinations, through professional registry and licensure examinations, to examinations designed to award equivalent credit, none of the available testing alternatives were really appropriate for program evaluation. Their content
only approximated the domain of any given institution’s curriculum and the results they produced usually provided insufficient detail to support improvement. But this did not prevent large numbers of institutions—especially those facing state mandates—from deploying them. One exception was the ACT College Outcomes Measures Project (COMP) examination (Forrest and Steele 1978). In many ways a harbinger, this examination was designed specifically to evaluate general education outcomes and to support group-level inferences about student learning. It also constructed general education outcomes in novel ways, emphasizing the application of knowledge in real-world situations and (in its long form) requiring authentic demonstrations of performance.

In the period 1986-89, the major testing organizations quickly filled the instrument gap with a range of new purpose-built group-level examinations aimed at program evaluation—all based on existing prototypes. Among the most prominent were the ACT Collegiate Academic Achievement Program (CAAP), the ETS Academic Profile, and a range of ETS Major Field Achievement Tests (MFAT). Student surveys provided another readily-available set of data-gathering tools, especially when they contained items on self-reported gain. While many institutions designed and administered their own surveys, published instruments were readily available including the CIRP Freshman and follow-up surveys, the College Student Experiences Questionnaire (CSEQ) and a range of questionnaires produced by organizations like ACT and NCHEMS.

The principal appeal of off-the-shelf tests and surveys in this period was their ready availability—a property enhanced when the first comprehensive catalogues of available instruments appeared (Smith, Bradley, and Draper 1994). Faced with a mandate demanding immediate results, most institutions felt they had little choice but to use such instruments, at
least in the short term. But there were also growing doubts about the wisdom of this approach (Heffernan, Hutchings and Marchese 1987), stimulating work on more authentic, faculty-made assessment approaches in the coming years.

- **Implementation.** A third challenge faced by early assessment practitioners was lack of institutional experience about how to carry out such an initiative. One question here was cost and, as a result, some of the first “how to” publications addressed financial issues (Ewell and Jones 1986). Others considered the organizational questions involved in establishing an assessment program (Ewell 1988). But absent real exemplars, the guidance provided by such publications was at best rudimentary. Enormous early reliance was therefore placed on the lessons that could be learned from the few documented cases available. Three such “early adopters” had considerable influence. The first was Alverno, whose “abilities-based” curriculum designed around performance assessments of every student was both inspiring and daunting (Alverno College Faculty 1979). A second was Northeast Missouri (now Truman) State University, which since 1973 had employed a range of nationally-normed examinations to help establish the “integrity” of its degrees (McClain 1984). A third was the University of Tennessee Knoxville, which under the stimulus of Tennessee’s performance funding scheme became the first major public university to develop a comprehensive multi-method system of program assessment (Banta 1985). These three cases were very different and provided a wide range of potential models. They were also unusually well documented, yielding some of the first concrete examples of assessment scholarship.

In the late 1980s, a “second wave” of documented cases emerged, including (among others) James Madison University, Kean College, Kings College, Ball State University, Miami-Dade Community College, and Sinclair Community College—many of which were responding to
new state mandates. To a field hungry for concrete information, these examples were extremely welcome. More subtly, they helped define a “standard” approach to implementing a campus-level program, which was widely imitated.

This founding period thus generated some enduring lines of assessment scholarship. One addressed concept development and building a coherent language. The purpose here was largely to stake out the territory—though much of this early literature was frankly hortatory, intended to persuade institutions to get started. A second line of work concerned tools and techniques and while all “forerunner” literatures were referenced here, strong reservations about standardized testing quickly emerged and persisted. A third strand comprised case studies of implementation, supplemented by a growing body of work addressing practical matters like organizational structures and faculty involvement. Finally, accountability remained a distinct topic for comment and investigation, looking primarily at state policy, but shifting later toward accreditation.

**Into the Mainstream**

By 1990, predictions that “assessment would quickly go away” seemed illusory. Most states had assessment mandates, though these varied in both substance and in the vigor with which they were enforced. Accrediting bodies, meanwhile, had grown in influence, in many cases replacing states as the primary external stimulus for institutional interest in assessment (Ewell 1993). Reflecting this shift, more and more private institutions established assessment programs. These external stimuli were largely responsible for a steady upward trend in the number of institutions reporting “involvement” with assessment. For example, in 1987 some 55% of institutions claimed they had established an assessment program on ACE’s annual *Campus Trends* survey.
By 1993, this proportion had risen to 98% (though the survey also suggested that most such efforts were only just getting started). Clearly, at least for administrators, assessment was now mainstream. But “entering the mainstream” meant more than just widespread reported use. It also implied consolidation of assessment’s position as a distinct and recognizable scholarship of practice.

- **An Emerging Modal Type.** As institutions scrambled to “implement assessment,” it was probably inevitable that they evolved similar approaches. And despite repeated admonitions to ground assessment in each institution’s distinctive mission and student clientele, they approached the task of implementation in very similar ways. As a first step, most formed committees to plan and oversee the work. Following widespread recommendations about the importance of faculty involvement, most comprised faculty drawn from multiple disciplines. But partly because the press to implement was so great, assessment committees rarely became a permanent feature of governance or of academic administration.

  The clear first task of these committees, moreover, was to develop an “assessment plan.” Often, such a product was explicitly required by an accreditor or state authority. Equally often, it was recommended by a consultant or by the burgeoning “how to” literature of practice (e.g., Nichols 1989). The resulting plans thus often had a somewhat formulaic quality. Most, for example, included a) an initial statement of principles, b) stated learning goals for general education and for each constituent discipline, c) a charge to departments to find or develop a suitable assessment method (frequently accompanied by a list of methods to be considered), and d) a schedule for data-collection and reporting. Implementing such plans, in turn, often involved the use of specially-funded “pilot” efforts by volunteer departments. Keeping track of implementation and reporting, moreover, often demanded use
of a tabular or matrix format (Banta 1996) and this too became a widespread feature of the “standard” approach. Methods, meanwhile, were healthily varied, including available standardized examinations, faculty-made tests, surveys and focus groups, and (increasingly, as the decade progressed) portfolios and work samples.

- **A Literature of Practice.** In assessment’s early days, the products of its scholarship comprised a fugitive literature of working papers, loosely-organized readings in *New Directions* sourcebooks, and conference presentations. But by the early 1990s, the foundations of a recognizable published literature could be discerned. Some of these works were by established scholars who summarized findings and provided methodological advice (Astin 1991, Pace 1990). Others tried to document assessment approaches in terms that practitioner audiences could readily understand (Erwin 1991, Ewell 1991). Still others continued the process of documenting institutional cases—of which there were now many—in standard or summary form (Banta and Associates 1993).

The establishment of the movement’s own journal, *Assessment Update*, in 1989 was also an important milestone in this period—providing relevant commentary on methods, emerging policies, institutional practices. As its editorial board envisioned, its contents were short, practical, and topical—providing the field with a single place to turn for ideas and examples. *Assessment Update*’s existence also provided an important alternative to established educational research journals for faculty-practitioners who wanted to publish. This supplemented the already-established role of *Change* magazine, which provided an early venue for assessment authors and continued to regularly print assessment-related essays (DeZure 2000). Through its Assessment Forum, moreover, AAHE issued a range of publications, building first upon conference presentations and continuing in a set of resource
guides (AAHE 1997). In strong contrast to fifteen years previously, assessment practitioners in 2000 thus had a significant body of literature to guide their efforts that included systematic guides to method and implementation (e.g., Palomba and Banta 1999), well-documented examples of campus practice (Banta, Lund, Black, and Oblander 1996), and comprehensive treatises integrating assessment with the broader transformation of teaching and learning (e.g., Mentkowski and Associates 2000).

- **Scholarly Gatherings and Support.** Initiated on a regular annual cycle in 1987, the AAHE Assessment Forum was by 1989 *the* conference for practitioners, providing a regular gathering-place for scholarly presentation and exchange. Sessions developed for the Forum required formal documentation and often ended up as publications. The Forum also maintained professional networks, promoted idea-sharing, and provided needed moral support and encouragement. The latter was especially important in assessment’s early years because there were few practitioners and they were isolated on individual campuses. Although the Forum remained the field’s premier conference, other gatherings quickly emerged. Some, like the Assessment Conference in Indianapolis (which actually began at the University of Tennessee, Knoxville), concentrated largely on orienting new practitioners. Others arose at the state level including (among others) the South Carolina Higher Education Assessment (SCHEA) Network, the Washington Assessment Group (WAG), and the Virginia Assessment Group (VAG)—often directly supported by state higher education agencies. Some of these state-level groups published regular newsletters updating members on state policy initiatives and allowing campuses to showcase their programs.

Funding support for assessment scholarship also became more accessible, primarily through the federal Fund for the Improvement of Postsecondary Education (FIPSE). In addition to
directly supporting assessment activities, FIPSE’s need for formal reports and evaluations helped stimulate the field’s growing inventory of published work (Cook 1989).

- A “Semi-Profession.” Although assessment remained largely a part-time activity, entering the mainstream also meant a rise in the number of permanent positions with assessment as a principal assignment. Position titles like assessment “coordinator” with formal job descriptions are now commonplace, usually located in Academic Affairs or merged with Institutional Research. The creation of such positions was in large measure a result of external pressure to put recognizable campus programs in place. Certainly such roles helped build badly-needed local capacity and infrastructure. But they in many cases also created real tensions about the ownership and benefits of the assessment process.

Early conversations, meanwhile, considered the advisability of creating a national professional organization for assessment similar to CAEL. A strong consensus emerged to maintain assessment as an “amateur” activity—undertaken by faculty themselves for the purpose of improving their own practice. Avoiding excessive professionalization was important because it promoted later linkages with the scholarship of teaching. But large and growing numbers of individuals on college and university campuses, often without conscious choice, have nevertheless adopted careers identified primarily with “assessment” as a distinguishable field.

For assessment as a whole, one clear result of entering the mainstream is an established community of practice that in some ways resembles an academic discipline. Among its earmarks are an identifiable and growing body of scholarship, a well-recognized conference circuit, and a number of “sub-disciplines” each with its own literature and leading personalities. Certainly this
is a significant achievement—far beyond what numerous early observers expected. But these very attributes have also decisively shaped, and in some ways limited, assessment’s impact on instruction and campus culture. Most campus assessment activities, for example, continue to be implemented as additions to curriculum, designed for purposes of program evaluation rather than being integral to teaching and learning. The fact that implementation so often centers on “doing assessment,” rather than improving practice through clear linkages to budget and pedagogy, moreover, can easily isolate the process from the everyday life of both faculty and administrators. Those doing assessment have evolved a remarkably varied and sophisticated set of tools and approaches and an effective semi-professional infrastructure to support what they do. But few faculty as yet practice assessment as a part of their everyday work. Firmly “in the mainstream” by the year 2000, assessment as a movement is thus still striving for the cultural shift its original proponents hoped for.

Episodes and Debates

Throughout its brief history, assessment has addressed a varied set of intellectual issues that have actively stimulated debate. Meanwhile, the movement went through several telling episodes that forced reaction and re-thinking. Encountering each prompted deeper understanding, though none have been entirely resolved. As a result, they continue to influence the course of this evolving scholarship.

- The “Ineffability” Debate. Perhaps the most basic debate that arises as faculties face assessment is the extent to which educational outcomes can be specified and “measured” at all. Indeed, a frequent early counter-argument was that any attempt to look at outcomes directly was both demeaning and doomed to failure (Benjamin AAU piece). Related
critiques noted that assessment’s principal vocabulary appeared confined to education and the social sciences—not always the most respected disciplines on any college campus. More pointedly both the rhetoric and implied methods advanced by the assessment movement have been frequently characterized as “positivist” and excessively mechanistic. Dissecting this classic complex of faculty reservations about assessment reveals some quite different underlying issues. Some are legitimately methodological, including appropriate reservations about the ability of off-the-shelf instruments and forced-choice methods to fully reflect collegiate learning, or fears about “teaching to the test.” Some are profoundly philosophical, based on a recognition that deep learning is always holistic, reflective, and socially-constructed. Still others are predominantly political, derived from faculty fears about loss of autonomy and creeping management control, as well as concerns about external intrusion into the curriculum.

What makes things complicated is that all three reasons are often bound up in a single sense of discomfort (e.g., Peters 1994). But the resulting debate about “ineffability” has proven helpful in deepening assessment scholarship. At one level, it has forced practitioners to sharpen the philosophical grounding of the movement—rooted in the tenets of scholarship and in the process of teaching and learning. It also re-emphasized that the evidence used by assessment must always rest upon a peer-based community of judgement (Mentkowski, Astin, Ewell, Moran, and Cross 1991, AAHE 1992). Finally, it forced explicit recognition of the fact that evidence is consistently constrained by the context in which it is generated (Mentkowski and Rogers 1988) and by the uses to which it is put (Messick 1988). Epistemological issues of this kind thus remain at the heart of the movement and remain
healthily and vigorously contested (Ewell 1989, Harris 2001). But protests based solely on principle or on politics have steadily diminished.

• **The “Value-Added” Debate.** The question of whether assessment’s primary focus should be placed upon documenting absolute levels of student attainment or on institutional contributions toward developing student abilities arose early (Ewell 1984). Certainly, reasons for centering attention on “talent development” were compelling. First, this approach was normatively appealing, and had the admirable property of leveling the playing field among different kinds of institutions. Assessing institutional “quality” in this way thus made more sense than using traditional markers like resources and reputation (Astin 1985). Ascertaining “net effects” also made good sense from a research point of view, recognizing that incoming student ability is the largest predictor of any outcome (Pascarella 1987, Pascarella and Terenzini 1991). Finally, some institutions were already practicing “value-added” approaches, and were finding them useful in demonstrating effectiveness (McClain and Krueger 1985).

But the classic approach to assessing learning gain—testing students on entry, then re-testing them at exit—posed perplexing conceptual issues and faced formidable methodological problems. Conceptually, it was argued, a pre-test was often simply silly because students had not yet been exposed to the subject being tested (Warren 1989). The terminology of “value-added,” moreover, suggested a mechanistic view of education in which students were viewed as “products” and learning merely additive. Actually determining growth, meanwhile, entailed multiplicative sources of measurement error and sometimes led to real misinterpretations of underlying phenomena (Hanson 1988, Baird 1988, Banta, et al. 1987). While active discussion of this topic diminished in the 1990s, it helped propel assessment
toward a useful synthesis. Most importantly, these discussions helped forge growing consensus that paths of student development should not be seen as linear and additive, but rather as organic and transformational. A methodological entailment of this growing consensus was longitudinal designs for assessment, capable of capturing large numbers of variables about both outcomes and experiences. Such longitudinal studies required an analytical model based on multivariate statistical control instead of simple “test-retest” approaches, and could be further enhanced by the use of qualitative methods like periodic interviews and focus groups. Finally, all agreed that for policy purposes, information about both levels of attainment and institutional contributions was needed.

- The TQM Episode. In the early 1990s higher education institutions began experimenting with Total Quality Management (TQM), a set of ideas and techniques borrowed directly from business, to help improve their administrative operations (Seymour 1991). Linking such notions with assessment was appealing because the two movements shared many attributes. Both began with a systemic approach to change and, indeed, viewed change itself as an imperative. Both emphasized the need to listen carefully to those the system was trying to serve—though the notion of students as “customers” immediately grated. Finally, both held that concrete information about performance was a critical part of a continuous cycle of planning and improvement. Recognizing such parallels, AAHE incorporated a track on TQM—quickly re-labeled Continuous Quality Improvement (CQI)—into its Assessment Forum, and issued a number of publications linking assessment and CQI (e.g., AAHE 1994).

But explicit attempts to fuse assessment and Total Quality were not successful, and after its initial flurry of activity, Total Quality has not fared well on campuses. Partly this was a matter of language. While assessment could ultimately adopt the discourse of scholarship,
Total Quality never shed its corporate flavor—especially in the eyes of skeptical faculty. Partly it was because the quality movement in business and industry itself had peaked. Yet much was synthesized by assessment—perhaps unconsciously—from this encounter. On the one hand, it reinforced “systems consciousness” and cemented the need to collect information about both outcomes and processes. Sometimes bitter “customer” discussions, meanwhile, helped underline the need to listen carefully to student voices, and to shift assessment’s perspective from “faculty teaching” toward students learning. Total Quality thus proved useful to assessment largely as a metaphor. At the same time, it taught object lessons about the risks of both alien language and the appearance of fad.

**The “National Assessment” Episode.** In 1990, the National Education Goals established the nation’s first objectives for collegiate learning. More specifically, they called for the development of valid and reliable assessments to track progress in critical thinking, communication, and problem-solving (National Education Goals Panel 1991). This action mirrored simultaneous and growing state interest in collegiate assessment, as a majority of the states had adopted assessment mandates for public colleges and universities by the mid-1990s. The Goals also signaled the beginning of a significant, though short-lived, period of aggressiveness by the U.S. Department of Education in the realm of postsecondary accountability, marked by such initiatives as the Student Right-to-Know Act and the State Postsecondary Review Entities (SPREs). Like the latter, the proposed national assessment never happened, though a major design effort to create it helped stimulate useful thinking about how a large-scale, authentic assessment of collegiate learning might actually be deployed (NCES 1992). Similar calls for a “NAEP for College” have periodically arisen, stimulated by both accountability demands and international comparisons (NCPPHE 2000).
To be sure, such episodes have had but little impact on the day-to-day practice of assessment on most college campuses. But together with their state-level counterparts, they provide a constant reminder that accountability was part of assessment’s birthright and is intimately entwined with its future. Continuing scholarship aimed at developing appropriate and timely responses to periodic accountability demands as they arise will thus always be needed.

Each of these issues illustrates how assessment discourse has grown in sophistication and has built a particular set of shared understandings. All remain centers of active debate. But few are now posed in the “black or white” terms in which they first arose and each helped stimulate improvements in methods and approach.

**Into the Future**

Social and educational movements, whatever their object, have one of two typical fates. Unsuccessful movements vanish after only a few years with little left behind. Successful ones disappear equally as “movements” because their core values become part of the dominant culture and their practices are fully institutionalized. So far, the assessment movement has experienced neither. One the one hand, levels of activity are unprecedented. The vast majority of institutions continue to report engagement with “assessment,” conference attendance is burgeoning, publications abound, and a growing body of practitioners see assessment as their primary professional practice. Yet at most institutions—and above all, for most individual faculty—assessment has not become a “culture of use” (Lopez 1997). The resulting paradox raises two questions, both highly relevant to the movement’s future. First, what sustained assessment for so long and what will continue to do so? And second, what prevented it from fulfilling its original promise and how might it ultimately achieve these ends? Answers to both questions, admittedly,
are speculative and uncertain. But as they play out, they will decisively affect the scholarship of assessment.

- **Why Didn’t Assessment Go Away?** In assessment’s first decade, the question of “when will it go away?” was frequently posed. This was largely because the movement was diagnosed by many as a typical “management fad,” like Total Quality or Management by Objectives (MBO), that would quickly run its course (Birnbaum 2000). Yet assessment has shown remarkable staying power and has undoubtedly attained a measure of permanence, at least in the form of a visible infrastructure. Several factors appear responsible for this phenomenon. Probably the most important is that external stakeholders will not let the matter drop. State interest is now stronger than ever, fueled by demand-driven needs to improve “learning productivity” and by burgeoning state efforts to implement standards-based education in K-12 education (Ewell 1997). Accreditation agencies, meanwhile, have grown increasingly vigorous in their demands that institutions examine learning outcomes, though they are also allowing institutions more flexibility in how they proceed (Eaton 2001). Market forces and the media are not only more powerful, but are also far more performance-conscious and data-hungry than they were two decades ago. Assessment has thus become an unavoidable condition of doing business: institutions can no more abandon assessment than they can do without a development office.

The last twenty years have also seen a revolution in undergraduate instruction. In part, this results from technology. In part, it reflects the impact of multiple other “movements” including writing across the curriculum, learning communities, problem-based learning, and service learning. Together, these forces are fundamentally altering the shape and content of undergraduate study. These changes are sustaining assessment in at least two ways. Most
immediately, new instructional approaches are forced to demonstrate their relative
effectiveness precisely because they are new. Assessment activities are therefore frequently
undertaken as an integral part of their implementation. More subtly, the very nature of these
new approaches shifts the focus of attention from teaching to learning. In some cases, for
instance, direct determination of mastery is integral to curricular design (O’Banion 1997). In
others, common rubrics for judging performance are required to ensure coherence in the
absence of more visible curricular structure (Walvoord and Anderson 1998). Assessment has
thus been sustained in part because it has become a necessary condition for undertaking
meaningful undergraduate reform—just as the authors of Involvement in Learning foresaw.

**Why “Broad but Not Deep?”** As important as assessment’s longevity, though, is the fact
that it has survived in a peculiar form. Most campuses are indeed “doing something” in
assessment. But the kinds of fundamental transformations in instruction that might have
resulted from examining systemic evidence of student learning have mostly not happened.
Instead, for the majority of institutions, assessment remains an “add-on”—done principally at
the behest of administration and sustained as a superstructure outside the traditional array of
academic activities and rewards. Reasons for this widespread condition, ironically, mirror
those that have sustained assessment for almost two decades. First, widespread and visible
external demands generally set the tone for initial engagement. Most campuses still do
“assessment” because somebody tells them to. Regardless of how the telling is done (and
external bodies have been more sensitive and flexible than is usually acknowledged),
responses risk being both reactive and mechanistic.

Like other efforts to accomplish meaningful undergraduate reform, moreover, assessment
must usually be implemented “across the grain” of deeply-embedded organizational
structures. Rewards for engaging in it remain scant for both institutions and individuals. So like similar activities not rooted in disciplines or departments like first-year experience programs or general education, assessment is frequently sustained as a separate activity, ensconced in an “office” and nurtured through special-purpose funding. Similarly, assessment results are rarely central to institutional planning and decisionmaking, even when undertaken outside the glare of public scrutiny. Partly this is because of continuing faculty fears about negative consequences. Ironically, it stems equally from faculty expectations of no consequences at all—that considerable effort will be expended gathering information that will never be used. Much of the appeal of the kinds of assessment activities that have been adopted on a widespread basis, like classroom research, is that the benefits of feedback are both immediate and apparent (Angelo and Cross 1993). At the institutional and program levels, the benefits of assessment have been far less immediately visible.

As this last observation suggests, two fundamental changes will be needed to transform assessment from a movement into a culture. One is at the level of teaching and learning, and requires shifting assessment’s conceptual paradigm from an evaluative stance that emphasizes “checking up on results” toward an emphasis on assuming active and collective responsibility for student attainment. Forces that might aid this conceptual transformation include the growing salience of ability-based credentials, which are fast becoming a way of life in many occupations and professions (Adelman 2000). Multi-institutional attendance patterns are meanwhile fueling demands to re-position articulation and transfer from course-based “seat time” toward performance-based attainment. Perhaps most important, reform efforts like writing across the curriculum and problem-based learning, together with technology, are forcing faculty to think far more concretely and collectively about learning outcomes and how to certify them.
A second needed transformation is at the level of academic administration, and requires evolving a largely top-down, “management-oriented” use of information in planning and decisionmaking toward a culture that more fully embodies the principles of a learning organization. Forces that might help this transformation are far less easy to identify, but include growing competition from non-university providers and insistent demands to create modes of instruction that are both efficient and effective.

Such developments, if they occur, will profoundly influence the scholarship of assessment. In its literature, there will likely be growing sophistication in discussions of methodology, capitalizing on emerging knowledge about how to forge consensual judgements of authentic performance (e.g., Walvoord and Anderson 1998, Mentkowski and Associates 2000) and about using technology to deliver complex, interactive problems. There will also be new demands for work on organizational transformation, in which assessment is addressed, but is fused with other systemic changes aimed at changing the environment for teaching and learning (e.g., Gardiner 1994, O’Bannion 1997, Harvey and Knight 1996). At a different level, assessment will gradually become an integral part of each faculty member’s reflective practice, documented through the scholarship of teaching (Shulman 1993, Hutchings 1996). And faculty will increasingly collaborate in this work, reflecting their growing assumption of collective responsibility for learning. Such developments are consistent with the tradition of robust, participatory, and practice-oriented scholarship already established by the assessment movement. If they emerge, they will constitute significant contributions to both theory and practice. More important, because assessment’s principal tenets will at last be embodied in the work of higher education on a day-to-day basis, its life as a “movement” may finally be over.
References


