National Institute for Learning Outcomes Assessment

The slow pace of change: Two histories of teaching and learning in U.S. colleges and universities

May 2021
Robert Hampel
Kevin R. Guidry
University of Delaware

How fortunate that Johns Hopkins University Press recently added these two books to its impressive offerings in higher education scholarship: Scott Gelber's (2020) *Grading the College: A History of Evaluating Teaching and Learning* and Jonathan Zimmerman's (2020) *The Amateur Hour: A History of College Teaching in America*. Each volume rests on exhaustive archival research and together they fill large gaps in the literature on the historical development of teaching and learning in U.S. colleges and universities

Evaluations of college teaching and learning developed slowly. Until the second half of the twentieth century, the instructor’s personality seemed to matter more than any other consideration. But there was no consensus on which admirable personal traits mattered most or which instructional strategies correlated with student learning. Vague terms like “good manners and general civility” or “a thorough gentleman” (Gelber, 2020, p. 28) appeared in letters of recommendation (so did comments on the candidate’s spouse). As one investigator concluded in 1932, “good teaching is a matter of men rather than of method” (Zimmerman, 2020, p. 105).

For Zimmerman, the heart of the story is in the word *amateur*. Most faculty and administrators did not think schools of education or departments of psychology could shed much light on the idiosyncratic craft of teaching. The research on elementary and secondary school teaching yielded no breakthroughs, the skeptics argued, so praise for the science of education was hyperbole. Why bother training graduate students to teach? Anyone smart enough to finish a good dissertation would be able to figure out what to do in class. It is no wonder that the periodic calls for a new doctoral degree for college faculty uninterested in research rarely won much support.

Both authors credit college newspapers for taking the lead on course evaluations. Questionnaires yielded both quantitative and qualitative profiles of dozens of courses, sometimes published in separate pamphlets. Some students used those documents to find easy courses but many used them to avoid dreary electives. Rarely did the faculty run their own parallel course evaluations (and when they did they were often optional and not used for promotion or annual review). By the late 1960s, pressure from student activists for mandatory course evaluations had increased to the point that most colleges accepted them as a low-cost, high-reward response to the tumultuous protests and student demands of the late 1960s. The greater importance of research for faculty careers remained in place, but at least there was an unprecedented acknowledgement that student views should be heard.

The new field of standardized testing sparked much excitement in the 1920s. Tests created by experts seemed less biased and more rigorous than teacher-made exams and many colleges began using external tests for admissions and advising. However, institutions rarely used them to evaluate teaching. Achievement tests seemed too focused on recall and too short of analysis and reasoning. Faculty also doubted if external exams could capture their specific goals and objectives. Concern was also raised that with norm referenced results, a mischievous invitation to compare their college to others could emerge. Comprehensive exams created by faculty fared better; the discussion about what to test often sparked valuable discussions of what and how to teach. But the time required to do the job well and the growing disagreement over what material every senior should know meant that many colleges backed away from comprehensives by the 1960s.

The greater importance of research for faculty careers remained in place, but at least there was an unprecedented acknowledgement that student views should be heard.

Gelber also traces the relative value and importance of grades as an indicator of learning. Although grades "privilege faculty authority and could represent complex qualitative judgments," (p. 92), there was significant work in the twentieth century to create alternative ways to measure learning. Following the development of standardized tests, rubrics and surveys were introduced in the middle of the twentieth century. Similar to the development of standardized tests, although some of this work was dependent on support and leadership from faculty researchers (e.g., NSSE supported by Indiana University's Center for Postsecondary Research), many projects were begun or supported by non-profit organizations like AAC&U, ACE, and ETS. Those developments, initiated outside of academia proper but very close to it, share similarities with the development of course evaluations by student newspapers.

The brief story of accreditation and how it intersects with assessment provides specific examples of how today's debates have been held many times in the past, often with the same unsatisfying resolutions. Gelber describes multiple movements toward and retreats from the inclusion of data about student learning in accreditation standards throughout the twentieth century. The AAUP's concerns in the 1950s that "the traditional measures of quality [are] losing credibility [particularly with state governments]" (p. 122) are familiar to readers seventy years later who still share those concerns.

In the final chapter of his book, Gelber tells us how he believes colleges should be evaluated. His brief recommendations focus on separating assessment into two separate strands, one focused on (external) accountability and the other focused on professional development for faculty. These recommendations echo observations and recommendations made by some assessment professionals who have been very challenged in meeting the multiple demands placed on them by disparate constituents. For Zimmerman, the application of what we already know is the challenge. Teaching can be evaluated, but when we do it, the tangible rewards for classroom excellence lag far behind the rewards for good research.

The major thread that runs through these histories, especially the history of assessment, is the continued involvement of faculty members. It is easy to place faculty members in the role of the "loyal opposition," reluctant to perform rigorous assessment of student learning especially in ways that are comparable across contexts. However, Gelber convincingly argues that much of the work in developing assessment tools and methods has been done by faculty members, particularly those who were scholars of measurement and evaluation or closely related disciplines. Zimmerman's history of teaching forcefully reminds us that much of the resistance to this work has revolved around fundamental disagreements about the nature of effective teaching and learning.

This may give us hope that the continued development of rigorous theories of teaching and learning supported by empirical evidence from multiple disciplines will pave the way for widespread acceptance of assessment tools and methods aligned with those theories.

Historians recognize that many of today's challenges, and the innovations developed to address them, are not new. Gelber presents examples of rubrics used to rate student work as early as the 1940s. During that same decade, he notes that scholars of evaluation began advocating for portfolios of student work that could be randomly sampled to "evaluate the attainment of goals at the course, program, or [institution] level" (93). In the 1920s and 1930s, colleges developed surveys asking students to self-report their learning to gauge the impact of new general education programs. Gelber describes efforts to measure "non-cognitive" factors dating back to the 1920s. That these challenges - and many purported solutions - have historical precedents underscores not only their lasting importance but their complexity.

Another theme throughout the history of teaching and learning in U.S. colleges and universities, particularly the assessment of learning, is the continued involvement of organizations other than colleges and universities. In some instances, these were membership organizations of colleges and universities e.g., ACE. In other instances, these organizations exist alongside and outside of academia e.g., ETS, the Carnegie Foundation for the Advancement of Teaching. These organizations, usually supported by philanthropic funds (or well-funded philanthropies themselves), sometimes worked with academia to advance a common agenda and sometimes worked counter to academia to advance their own agenda. But their influence was and remains incredibly important.

This final thread in Gelber's history of assessment - the slow development of rigorous assessment tools and methods - has an obvious parallel in Zimmerman's history of teaching in U.S. colleges and universities. Gelber's history can be read as the development of assessment as a profession or one carried out by well-informed and experienced professionals. Zimmerman's history of teaching is largely the story of teaching carried out by amateurs. If that is finally changing - if teaching experience and evidence of effectiveness is highly valued in the extremely competitive faculty job market, if institutions and organizations increasingly provide effective coaching in teaching and learning for graduate students and faculty, and if more faculty understand how people learn-- then we can look forward to more faculty appreciation of the value of assessment.

The histories of teaching and learning presented by Gelber and Zimmerman make it clear that changes in these areas occur slowly. Forces within and outside the academy have often pressed for changes - improvement, modernization, professionalization, etc. - but have rarely succeeded in winning the trust and confidence of faculty members who have traditionally controlled teaching. Assessment experts in the early 21st century are left wondering what, if anything, has changed. Will our efforts be remembered as yet another in a long series of well-intentioned but ultimately unsuccessful attempts to change institutions that successfully resist changes, especially changes perceived as coming from outside of academia? Or have enough things changed - empirical knowledge about teaching and learning, expectations of students and parents, increasingly well understood tools and methods, and intensive financial pressures - to position us differently than our forebearers to make long-lasting and effective improvements to teaching and learning?

Or have enough things changed - empirical knowledge about teaching and learning, expectations of students and parents, increasingly well understood tools and methods, and intensive financial pressures - to position us differently than our forebearers to make long-lasting and effective improvements to teaching and learning?

Please Cite As:

Hampel, R., & Guidry, K. R. (2021, May). *The slow pace of change: Two histories of teaching and learning in U.S. colleges and universities.* Urbana, IL: University of Illinois and Indiana University, National Institute for Learning Outcomes Assessment.

Follow us on social media:

@NILOA_web

@LearningOutcomesAssessment

Sign up to receive our monthly NILOA Newsletter and stay up to date with our research and publications.