From Steven Brint, *Two Cheers for Higher Education: Why American Universities Are Stronger than Ever – and How to Meet the Challenges They Face.* Princeton: Princeton University Press. Scheduled for publication: Fall 2018.

Chapter 8
Focus on the Classroom

In the next two chapters I take up the principal threats to the continued expansion and prominence of U.S. colleges and universities. I devote the current chapter to the weakness that strikes at the core of academe's objectives and makes the triumph of online alternatives more plausible. The issue has been diplomatically described as "underachievement" in undergraduate education (Bok 2006), but it could be described equally well as the failure to inculcate professional standards and expectations for college teachers. I will analyze why the disparate efforts during the period to reform undergraduate teaching and to make colleges accountable for student learning failed to transform college classrooms. I will also show why the new sciences of learning have potential to create the more powerful learning environments that earlier reformers failed to produce. In chapter 9, I discuss other major challenges to the U.S. higher education system: rising costs, online competition, and controversies over permissible speech.

I do not want to overstate the threat these challenges pose. The top one hundred or so research universities and the top thirty or so liberal arts colleges are in no danger from those who hope to disrupt the system; they are strong enough to weather any conceivable threat. Some other institutions, including "brand-name" religiously affiliated colleges and universities, such as Brigham Young University and the University of Notre Dame, can count on supporters who will not allow them to decline, and so can other campuses dear to alums. But a sizable proportion of poorly funded four-year public colleges and universities are at risk, not necessarily for closing their doors but rather for becoming industrialized online course distributors stripped of a campus culture and intellectual aspirations. At risk also are hundreds of low-enrollment, low-endowment

private colleges.

Academic skill development is no longer considered the only important aim of undergraduate education, and some do not consider it the most important aim. Some social scientists argue that the primary purpose of college is to provide a grounds for students to develop social network ties (Stevens, Armstrong, and Arum 2008) or to socialize them into the "posture of actorhood," (Meyer 2008). Others focus on what they see as a rough equality among the multiple forms of engagement that undergraduates can pursue – academic, co-curricular, or civic engagement (Douglass and Thomson 2017). These ways of seeing the undergraduate experience reflect real changes in what students and their parents want out of college. Students and their parents tend to express greater interest in what happens outside than what happens inside the classroom (Clotfelter 2017, 179-85). One can take a student-led tour of even the most prestigious Ivy League institutions, as I have done, and hear enthusiastically delivered testimonials to the opportunities that lie ahead in student clubs and organizations, but barely a word about what happens in classrooms.

In previous work I have argued that colleges' responsibilities for student development should ideally go beyond the classroom to include contributions to their personal development, interpersonal skill development, opportunities for civic engagement, and preparation for economically secure employment (see Brint 2015b). At the same time, it seems implausible to me that colleges would want to hang their hats on these alternative engagement pegs. Other organizations from Toastmaster's to the Red Cross could perhaps do as well in non-academic arenas of student development. But we do need colleges and universities to help students to analyze and question complex texts, compose persuasive arguments, understand scientific concepts, sift evidence to come to conclusions, weigh alternative explanations, and for many

2

other cognitive skills that educators have reason to expect will lead to clearer and more effective thought -- and action.

It is a good question whether students' fates really depend on the development of their academic skills. I have already observed that many employers say they value qualities such as deference to authority and cooperation with peers as much as they value cognitive skills. Nevertheless, most of the evidence suggests that cognitive skills and economic well-being are related. A 2012 OECD program tested young adults (ages 16-34) on literacy, numeracy, and computer-aided problem solving skills. Scores at a proficiency level were strongly associated with higher wages for young adults, with an average 22 percent gain in income for every standard deviation change in combined literacy and numeracy skills (OECD 2016). (Measures of good health were also strongly related to skills.) Unfortunately, these data do not control for pre-existing background differences, and socio-economic background status showed a strong impact on U.S. scores. Even so, they are compatible at the individual level with national level findings on rates of GDP growth. Countries in which students score high on the OECD's Programme for International Student Assessment (PISA) tests at age 15 tend to have significantly stronger growth rates in later years than countries in which students on average score low, even after controls are introduced to account for potentially confounding influences on GDP growth (Hanushek and Woessmann 2011, 160-87).

Not all successful countries require intensive periods of study during the college years. Japan, for example, is well known to demand little of college students who pass a pleasant, nearly study-free period between the intense competition of high school preparation and business employment. But unlike Japanese students most U.S. college students cannot claim that rigorous secondary school experiences have prepared them for experiences that require high levels of
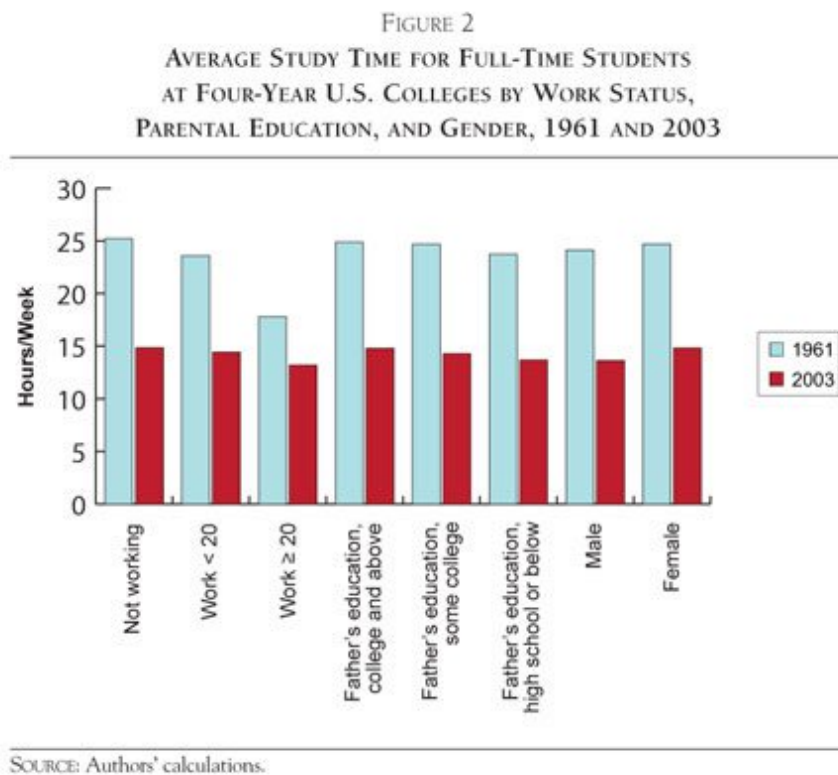
literacy, numeracy, or problem-solving skills after college. To the extent that higher education institutions are failing to contribute sufficiently to students' cognitive development, they can be considered to be failing in a fundamental way.

Students' Declining Commitment to Study

Concerns about undergraduate students' educational experiences go back many decades, but strong empirical evidence of the shortcomings of undergraduate education began to emerge only in the 2000s. The 2003 National Assessment of Adult Literacy showed that only 30 percent of college graduates could accurately interpret two competing editorials or make accurate inferences from a graph relating age, exercise, and blood pressure (Kutner et al. 2007).[1] A study of the analytical and critical thinking skills of 2,400 college students by Richard Arum and Josipa Roksa (2011) concluded that nearly half of the students studied made no significant gains in their reasoning capacity between freshman and the middle of sophomore year. A subsequent follow-up showed that more than one-third of these students made no statistically significant gains between freshman and the end of their senior year (Arum, Roksa, and Cho 2012). The study can be criticized; students had no clear incentives to put out maximum effort on the test. But the study has been replicated using alternative tests of analytical and critical thinking, with remarkably similar results (Pascarella et al. 2011). The aforementioned OECD assessment of young adults' skills also yielded an avalanche of disturbing information about the state of Americans' skills relative to those of young adults in the rest of the developed world. Americans ranked number 16 of 22 nations on the literacy test and they tied for last place on both the numeracy and computer-aided problem solving tests. Nor did the country's top achievers rank high internationally. The 90th percentile of scorers ranked sixteenth compared to their

international counterparts; those with bachelor's degrees, twentieth; and holders of post-graduate degrees, nineteenth

Beyond these disappointing test results, the economists Philip Babcock and Mindy Marks (2010) found a secular decline in study time from the early 1960s to the mid-2000s, with students reporting about half as much study in the later period as in the earlier period. In the early 1960s, students reported studying and attending class approximately forty hours per week. By the mid-2000s, study time had fallen to approximately twenty-seven hours per week. Babcock and Marks found comparable declines at every selectivity level, in every major, and among every demographic group, albeit from markedly different starting points (see figure 8.1).

FIGURE 2
AVERAGE STUDY TIME FOR FULL-TIME STUDENTS
AT FOUR-YEAR U.S. COLLEGES BY WORK STATUS,
PARENTAL EDUCATION, AND GENDER, 1961 AND 2003

SOURCE: Authors' calculations.

These studies suggest that either teachers are not requiring much of their students or students are finding ways to minimize the time they spend on class with the acquiescence of their teachers. My own work with Allison Cantwell on time use among University of California

students found that one-fifth of those responding to a census survey fielded in 2012 said they spent less than eighteen hours a week in class or studying for class. Another one-fifth reported that they read half or less of assigned reading for their classes. And one-quarter reported rarely, if ever, participating in class or being in touch with their instructors (Brint and Cantwell 2014). These findings were alarming, in part, because we knew that more conscientious students are the ones more likely fill out a lengthy survey.

The failure of instructors and administrators to attend to weaknesses in undergraduate teaching and learning can be considered the Achilles' heel of U.S. universities. I will focus on three persistent weaknesses: (1) teaching that encourages student passivity, (2) the limited accountability instructors expect of students, and (3) teaching for rote memorization rather than for understanding and mastery. These weaknesses are often interrelated as inadequately trained instructors dominate classroom air time without providing opportunities for students to interact with one another around course materials, fail to build in mechanisms to ensure that students are prepared for and paying attention in class, and orient their courses toward an emphasis on what students need to know for tests rather than a mastery-level understanding of subject matter.

These problems do not, of course, fill up every corner of higher education. Expert and impassioned teachers such as Columbia English professor Andrew Delbanco have described the epiphanies that occur in discussions that "envelop the mind in multiple perspectives" and, quoting William James, lead to the "ideal vanishing point towards which we imagine that all our temporary truths will someday converge" (2012, 60). The opportunities for creating challenging, intensely engaging courses are at their maximum in small seminars in which students encounter skilled teachers and are inspired to be prepared—or feel they must be prepared for fear of disappointing their instructors and fellow students. Such experiences are the norm in doctoral

education. The educational appeal of the country's best liberal arts colleges and private research universities is that they can provide a comparatively high number of these educational opportunities. But virtually every campus can point to at least a few dozen classes that approach or meet this educational ideal.

Dimensions of Instructional Effort

It is useful to think of college teaching as involving at least three dimensions of instructional effort. One is to adequately convey subject matter. Another is to help students to make the transition from inadequate to adequate understandings of subject matter. The third is to provide sufficient rigor in assignments to produce growth. One cannot expect that rigor will look the same at elite colleges as at broad access colleges, but every instructor can push his or her students at least a little beyond their comfort levels.

There are good reasons for doing so. Many years ago, Karen Miller and her colleagues (1985) measured the substantive complexity of schoolwork by examining the intellectual difficulty of courses, the length of time it took to complete work for them, and the complexity of out-of-class projects. They found that students who experienced greater substantive complexity in their work and less detailed supervision by teachers scored higher on "educational self-direction." These students were required to use more initiative, engage in deeper thought, and exercise more independent judgment. Unfortunately, the college-for-all system provides these deep-learning opportunities for only a minority of students.

University administrators can take their share of the responsibility for the limited learning that so often occurs in college classrooms. It is, after all, the institutional decision makers who have reduced the ratio of instructional staff per student and increased the ratio of administrative staff per student. They are also the ones responsible for bringing so many adjunct instructors into

classrooms without providing them with adequate pay, working conditions, or supervision. The studies of the economist John Bound and his colleagues underscore the extent to which gaps between access and graduation are the result of institutional factors such as large class sizes and insufficient feedback on course work (see, e.g., Bound, Lovenheim, and Turner 2010). But the problems go beyond the "graduation gap"; the "throughput" models used to measure efficiency by charting rates of graduation and years to graduation do not capture anything about whether graduating students are learning skills of value while they are on campus. They are about input and output and nothing in between.

It is here that we see a fundamental tension between the university's value commitments and its financial drivers. Its desires to add value to students through effective teaching practices are consistently challenged by its financial incentives to maximize and optimize resources. The "competing logics" framework I developed in chapter 1 suggests that this tension can also be interpreted as a conflict between the logic of academic professionalism, which should embrace instructional quality (though it does not always), and the market logic of enrollment expansion. One reason for believing that undergraduate education will continue to underachieve is that institutional survival and well-being is inevitably more important to university administrators than difficult-to-measure values like educational quality. It follows that practices associated with student satisfaction will take precedence over practices not associated with student satisfaction.

But these priorities do not necessarily create insuperable obstacles to the improvement of undergraduate teaching and learning. A robust research literature has developed on techniques that work to aid student learning. A much stronger teaching and learning environment is consequently on the horizon, if colleges and university faculty members have the will to adopt approaches that lead to better results for more students. Many are already doing so. Campuses

such as Alverno College in Wisconsin, the University of Maryland-Baltimore County, and the University of Texas-El Paso, have gained reputations for teaching excellence over the years, either by focusing on rigorous assessments of student learning (the Alverno case) or through the extra efforts they have made to build scholars and scientists from among socioeconomically disadvantaged students (the UMBC and UTEP cases). University administrators can do their part by providing the resources to help faculty members become expert teachers and to make potentially transformative experiences like undergraduate research, study abroad, and skills-based learning communities available to more students.

The Contradictions of Postwar Academe

By the mid-1960s, the trend toward populating academe with professional researchers was so noticeable that Christopher Jencks and David Riesman coined the term "the academic revolution" to mark what they assumed would be a permanent turning point in the shift of the profession from teaching to research (Jencks and Riesman 1968). For research university professors, the requirement to meet the exacting standards of colleagues evaluating articles and books warranted careful training; half-awake, half-interested undergraduates sitting in the back rows of large lecture halls were another matter. In graduate training programs of the period, students were not required to demonstrate skills in pedagogy during their studies for the PhD, or understanding of the relation between types of pedagogy and subject matter content, or understanding of the aims or purposes of education. Rather, those who were not fortunate or promising enough to obtain research assistantships were thrown into teaching discussion sections without preparation, under the presumption that anyone smart enough to be in graduate school was smart enough to run a discussion section. For most would-be professors, teaching was an amateur activity, performed with limited regard to effectiveness, by people whose real training

9

was for something else entirely.

Many observers within the university welcomed this era of the research-centered professoriate. For Clark Kerr, the new multiversity served the nation by providing greater access, scientific and scholarly progress, and expert advice to every constituency in its state and region. But, Kerr acknowledged, undergraduate teaching suffered:

> There seems to be a "point of no return" after which research, consulting, [and] graduate instruction become so absorbing that faculty efforts can no longer be concentrated on undergraduate instruction as they once were. (1963, 65)

Kerr provided no solution to the "cruel paradox that a superior faculty results in an inferior concern for undergraduate teaching," though he hoped that an escape from the paradox could eventually be found (ibid.).

More astringent critics, like Columbia provost Jacques Barzun, pointed out the injustice of shortchanging undergraduate students:

> The student … is conscious (that his teachers) subject him to cavalier treatment … unpunctual, slipshod in marking papers, ill-prepared in lecture, careless about assignments. … To put it another way, the student sees and resents the fact that teaching is no longer the central concern of the university. … After making all due exceptions (for there are still thousands of devoted teachers and vigilant college heads), the students' complaint is justified. The great shift to research after 1945 would alone modify the university atmosphere sufficiently to warrant the impression of neglect, supported as it is by the reality of "publish or perish." (1968, 69)

Although Barzun and others (see, e.g., Schaar and Wolin 1965) expected a student uprising against desultory and negligent undergraduate teaching, these hopes were quickly disappointed. Instead, an ethic of consumerism emerged. This ethic reflected the growth of mass higher education, which brought many more ill-prepared and non-academically oriented students to campus. Moreover, students now had the power, in the form of student evaluations, to register their desires in ways that affected teachers' expectations. The use of student evaluations of teaching became widespread in the 1970s (Riesman 1980). At large state universities, these

10

forms became the primary method for evaluating performance in the classroom. Feedback from students about "too much reading" and "unfair tests" encouraged instructors to lower their expectations of student work in the hope of retaining high evaluations and in response to a declining academic ethos among students (Everett 1977; Grubb 1996; Johnson 2003; Riesman 1980).

Two Reform Movements Emerge

It is against this backdrop that two movements emerged during the period to address weaknesses in undergraduate education.[2] The limited success of these movements demonstrates how difficult it is to create meaningful change in undergraduate teaching and learning in a heterogeneous system lacking any semblance of coordination or any widely accepted standards for what constitutes effective instruction.

Sharing a critical stance toward the condition of undergraduate teaching and learning, the two movements otherwise shared little in common: *the teaching reform movement* worked on a doctrine for the improvement of teaching effectiveness, while *the outcomes assessment movement* focused on requiring institutions to assess how much students were learning in their classes. The higher education policy analyst Peter Ewell described the two movements as they emerged near the beginning of the period:

> Two antithetical "ideologies" … arose almost simultaneously in higher education discourse. The first came from inside the academy. … Its tenets were most clearly stated in an influential national report, *Involvement in Learning* [1984] … which argued that breakthrough improvements in undergraduate education could be achieved by establishing high expectations, deploying active and engaging pedagogies, and providing feedback about performance. … The second ideology had roots outside the academy based on strong state interest in pursuing [testing-based] educational reform. … Its tenets were embodied in a high visibility report by the National Governors Association, *A Time for Results* [1986]. … The report argued that colleges and universities should be held accountable for establishing clear standards for performance with respect to student learning and that the results of student assessments should be publicly reported and coupled with consequential actions. (2005, 107).

It is clear that one strand of the teaching reform movement gained considerable ground during the period. Networks of teaching practitioners succeeded in disseminating selected principles of what I will call *the new progressivism*—specifically, those principles promoting student-centered learning, civic and community engagement, and sensitivity to the interests of diverse learners. In this way, they contributed to improvements in students' interactive engagement in the classroom, but not necessarily for their accountability for learning or their deeper understanding of course materials. By contrast, and perhaps surprisingly, the outcomes assessment movement failed to transform practice, even as it put learning outcomes more emphatically on the policy agenda.

The Teaching Reform Movement

The principal agents of the teaching reform movement were the great philanthropic foundations and foundation-sponsored advocacy organizations, such as the American Association of Colleges and Universities (AAC&U) and the Carnegie Foundation for the Advancement of Teaching (CFAT). Supported by the foundations, the principles of good teaching codified by leading educators of the period advocated active learning experiences, commitments to diversity and civic engagement, *and* challenging academic standards. Challenging academic standards were repeatedly extolled, based on evidence that learning improvement requires meeting high expectations. However, this advocacy of challenging academic standards proved to be no match for the consumerism and utilitarianism that came to dominate college student academic life. The trajectory of the teaching reform movement consequently mirrored the pattern of K–12 progressive education in the early twentieth century, when followers of John Dewey, such as William Heard Kilpatrick, deemphasized his mentor's insistence on rigor and frequent assessment and highlighted the student-centered, active learning, and community engagement

themes in his work (Cremin 1961, 328; Labaree 2005).

"Good Practices" in College Teaching

The National Institute of Education's influential *Involvement in Learning* (1984) signaled both the growing importance of effective teaching and the challenges facing faculty in a system of mass higher education. This document, heavily influenced by the thinking of UCLA higher education professor Alexander W. Astin, advocated movement away from the standard lecture format so that students could become inquirers—producers, as well as consumers, of knowledge. Following the lead of progressive educators, the report recommended the introduction of active modes of learning, such as faculty research projects and classes held in the field; internships and other forms of carefully monitored experiential learning; small discussion groups; in-class presentations and debates; and individual learning projects and supervised independent study. It also advocated timely feedback and more rigorous standards for evaluating student performance (National Institute of Education 1984, 27–28).

Arthur W. Chickering and Zelda Gamson's "Seven Principles for Good Practice in Undergraduate Teaching" promoted a similar set of recommendations. Their easy-to-remember principles became a touchstone for reformers and formed a basis for subsequent national surveys of student engagement. In a pithy opening sentence, Chickering and Gamson identified the complaints brought against colleges and universities in an age of mass higher education: "Apathetic students, illiterate graduates, incompetent teaching, impersonal campuses—so rolls the drumfire of criticism" (1987, 3). As in the case of the National Institute of Education report, the seven principles offered something for both progressives (frequent faculty-student contact, collaborative and active learning experiences, and respect for the variety of students' talents and ways of learning) and traditionalists (focus on time spent on task, prompt feedback, and high

13

expectations for performance).

Scholarship Reconsidered

Ideologies provide blueprints for action, and by the end of the 1980s organizational changes had created the conditions for an ideological shift—from the research-centered hierarchy of the "academic revolution" to something new reflecting the variety of institutional missions found in U.S. higher education. That new ideology was formulated in Ernest L. Boyer's *Scholarship Reconsidered* (1990). As president of the Carnegie Foundation for the Advancement of Teaching, Boyer was well positioned to affect change in institutional practices.

Boyer's underlying goal was to install a confederation of interests in the place of academic hierarchy. To do so, he identified four legitimate forms of academic life: the scholarships of discovery, integration, application, and teaching. The use of the venerable term "scholarship" united academe under the idea of studiousness and learning rather than research and teaching. Boyer explicitly hoped to end debates about the relative value of research and teaching. "The most important obligation now confronting the nation's colleges and universities," he wrote, "is to break out of the tired old teaching versus research debate and define, in more creative ways, what it means to be a scholar. It's time [for the profession] to recognize the full range of faculty talent and the great diversity of functions higher education must perform" (xii).

The critical innovation in Boyer's work was the integration of teachers as equal partners in the confederation of scholars. Before Boyer, one rarely thought of teaching as scholarship, only as reflecting knowledge of scholarship. Although the term "scholarship" suggests the possibility of professionalizing the teaching function, for Boyer it remained the province of the inspired amateur, albeit one who thought deeply about subject matter and reflected often on the

14

effectiveness of her practice. Yet the very naming of teaching as a form of scholarship encouraged steps in the direction Boyer himself initially failed to anticipate, toward research on teaching effectiveness that could be used to inform practice.

Boyer's work undoubtedly contributed to raising the stature of teaching as an object of concern and as a central identity for academics. In a national survey of postsecondary faculty conducted five years after the Boyer report, more than three-quarters identified teaching as the most important activity in their professional lives (Schuster and Finkelstein 2006, 87). The faculty as a whole reported that 60 percent of its work time was spent on average on teaching-related activities, as compared to 15 percent on research (ibid., 88). Only the natural and social sciences and engineering showed any reapportionment of effort in the direction of research (ibid., 91). In addition, institutions more often required evidence of "teaching excellence" in applications for positions following the Boyer report; such evidence was required in 60 percent of advertisements placed in the *Chronicle of Higher Education* (Meizlish and Kaplan 2008). These requirements grew at research universities, as much as baccalaureate- and master's-granting institutions, and particularly in the arts and humanities.

Ernest Boyer hoped to maintain scholarship at the center of the profession. Yet the American College Faculty surveys suggest that the centrality of scholarly contributions slowly eroded in the face of the participatory practices and eleemosynary goals of professors. Among full-time faculty in public doctoral-granting universities, interest in becoming an authority in one's field declined by 10 percent between 1989 and 2004, before increasing a bit in 2007. Interest in obtaining recognition from colleagues for scholarly achievements showed a similar rate of decline. Indeed, obtaining recognition from colleagues for one's scholarly contributions was no longer a goal held by a majority of faculty in public master's-granting institutions, even

as helping others remained a primary goal. American college faculty outside of private universities were more likely to say in 2007 that helping others in difficulty was a more important goal than becoming an authority in one's field or obtaining recognition from colleagues for scholarly contributions (DeAngelo et al. 2007).

These data suggest that support for teaching did not preserve scholarship as the unifying feature of the academic profession but rather that college teaching in many institutions and fields was transformed from more of a scholarly profession into more of a helping profession. Academics who did not conduct much research began to develop their own norms of practice emphasizing active learning experiences, inclusive pedagogies, and social service goals. Boyer expected pluralism to strengthen the usefulness and unity of the profession. But the scholarship of teaching was not an intrinsic feature of the new progressivism that emerged out of the teaching reform movement.

The New Progressivism

A new progressivism emerged as the leading outcome of these reform movements. It focused more on student engagement and inclusive practices than on student performance. Practitioners found engagement activities easier to implement than challenging assignments—and less likely to create student discontent. For all of its successes, the new progressivism raised a central question: Can a strong academic profession be built around a teaching identity focused more on student engagement and social amelioration than on a commitment to the practices of scholarship and research?

The American Association of Colleges and Universities (AAC&U) is the organization most responsible for extending the good practices literature to include attention to the classroom

climate for disadvantaged groups..[3] As I discussed in chapter 4, the initial work of the AAC&U was to create a warmer campus and classroom climate for women and minorities, first by publicizing their marginalization and then by connecting their full incorporation into the democratic project. This early work culminated in the American Commitments initiative (1993–2001), funded by the Ford Foundation, the Hewlett Foundation, and the National Endowment for the Humanities. The connection between diversity and democracy provided a signal theme for this work. AAC&U drew on familiar images of pluralism but with a new twist: "Higher education," it wrote, "can nurture Americans' commitment and capacity to create a society in which democratic aspirations become democratic justice. Diversity proves a means of forging deeper civic unity" (Beckham 2000, 2). This conceptual link between diversity and democracy brought diversity thoroughly into the mainstream of liberal education, while updating the Deweyan tradition to incorporate the race- and gender-conscious movements on campus.

The National Survey of Student Engagement (NSSE) represented another powerful force in the institutionalization of the new progressivism. Led by George D. Kuh, a professor of higher education at Indiana University, NSSE was launched with Pew Foundation funding in 2000. NSSE built on decades of research by Kuh and his colleague Robert Pace on the College Survey of Educational Quality (CSEQ) (Kuh 2009). This work closely paralleled the precepts of *Involvement in Learning*. Conceived in part as an alternative to resources- and reputation-based college rankings of *U.S. News and World Report*, NSSE intended to measure more accurately the actual quality of undergraduate students' educational experiences. The five NSSE benchmarks, each addressed through scaling-related questions, probed levels of (1) student-faculty contact, (2) active and collaborative learning, (3) academic challenges, (4) educational enrichment activities, and (5) institutional climates conducive to learning.

In its inaugural year, NSSE was administered at more than 270 institutions; this number grew to more than 600 annually by the end of the decade (www.nsse.iub.edu). Institutions were soon comparing their engagement scores on the five key dimensions to national norms and norms for institutions of their type. NSSE generated an impressive number of reports detailing the distribution and consequences of engagement experiences, and it also championed case analyses of institutions that showed exceptional effectiveness in the production of engaged learning environments (Kuh et al. 2005). It developed a checklist of "high-impact" practices that found their way into the curriculum of hundreds of colleges. These included freshman seminars, senior capstone experiences, study abroad, undergraduate research, service learning experiences, and internships (Kuh 2008).

However, NSSE measured engagement, not learning,[4] and, although many college educators assumed that higher levels of engagement should register more or less directly in improved learning outcomes, empirical efforts to demonstrate this proposition were disappointing. Student scores on NSSE scales were, for example, only very weakly associated with scores on the Collegiate Learning Assessment (CLA), a test of analytical and critical thinking, and most factors failed to reach statistical significance once students' prior academic records (grade point average and SAT scores) were statistically controlled (Carini, Kuh, and Klein 2006). Other studies showed that high grades were common in humanities and social science courses in which the culture of engagement emphasized participation, interaction, and active learning experiences and were less common in the natural sciences and engineering where engagement typically meant long hours of study, with groups of peers, to master demanding quantitative material (Brint, Cantwell, and Hanneman 2008; Johnson 2003).

Changes in Classroom Practices. Classroom practices changed dramatically in the

18

direction advocated by the new progressives, even as the more traditional-sounding parts of the Astin-Chickering teaching reform message, those focusing on high expectations and challenging assignments, were lost or ignored. Here the best data come from the triannual studies of American faculty from the Higher Education Research Institute (HERI). From the late 1980s through the mid-2000s, extensive lecturing showed a marked decline as a teaching method, even in public research universities, and cooperative (small-group) learning opportunities a corresponding increase. Full-time college faculty increasingly said they were bringing their students into field settings; asking them to demonstrate their knowledge in front of class through oral presentations; relying on reflective writing and journaling; using real-life problems to illustrate lessons; and putting student-centered inquiry, rather than recitation of facts and concepts, at the center of their teaching work (Astin, Dey, and Korn 1991; DeAngelo et al. 2007; Dey et al. 1993; Lindholm et al. 2002; Lindholm et al. 2005; Sax et al. 1996; Sax et al. 1999).

These changes went together with an expanded conception of the goals of undergraduate education. Consistent with principles of the new progressivism, the American College Faculty studies also showed sharp increases in the centrality of social goals as well: reaching out to surrounding communities through community-based research; teaching appreciation of multicultural diversity; and interest in using undergraduate education as a vehicle for promoting social change. Just as the twentieth-century progressives socialized their ideals of citizenship through the schools, so too do college faculty now overwhelmingly endorse the goals of diversity and community engagement. The main proponents of these changes were younger and female faculty members (DeAngelo et al. 2007, 5, 9, 11), suggesting that these trends are likely to continue as older faculty retire and college teaching faculties become increasingly populated by those brought up in the norms of the new progressivism.

19

Engagement versus Learning. The new progressives assumed that higher levels of engagement would lead more or less automatically to better student performance, as enhanced engagement triggered enhanced motivation to study. But the evidence indicates that challenging academic work and mechanisms to ensure accountability for learning are also very important features of instruction.[5] Looking at a sample of 2,400 students who took the CLA at the beginning of their freshman year and the middle of their sophomore year, Arum and Roksa (2011) found that students had improved their critical thinking, complex reasoning, and writings skills, as measured by the CLA performance task, by only 0.18 standard deviations, or an average seven percentile gain. Forty-five percent of students showed no change in their CLA scores. Arum and Roksa concluded that students' completion of three semesters of college had made a "barely noticeable" impact on the higher-level cognitive skills tested by CLA. They attributed the primary reason for this poor performance to the unwillingness of college instructors to give challenging assignments, which they measured, somewhat arbitrarily, as fewer than forty pages of course reading in a week and/or no assignment of papers of fifteen or more pages during the semester.

Trend data from NSSE confirm their conclusion. They showed that many active and collaborative learning activities grew more popular over time, while challenging requirements, such as the amount of time students spend studying per week and the number of twenty-page papers they wrote, remained static or declined (NSSE 2000, 2008). In the 2008 NSSE report, nearly two-thirds of seniors in NSSE sample institutions said they studied fifteen or fewer hours per week, and half said they had never written a paper of twenty pages or longer (NSSE 2008). In both cases, challenging requirements were less common in 2008 than those found eight years earlier.

One reason for this one-sided adoption of "good practices" research is clear: under-prepared and unmotivated students tended to penalize demanding teachers in their student evaluations (Babcock 2010), leading many instructors to worry that raising expectations would have negative consequences for their careers. The reliance on student evaluations of teaching as the sole measure of good teaching practice cemented student consumerism as an effective force in the classroom and one that tended to lower rather than raise expectations for learning, thereby undermining important planks in the good-practices platform.

The Triumph of Consumerism. The triumph of student consumerism is evident in these findings. Many students have effectively resisted professorial demands for higher levels of effort by simply refusing to engage their studies at a deep level. Ethnographic studies indicate students have relied on posted lecture notes, the prevalence of relatively easy courses to fill out their schedules, and teachers' openness to negotiations concerning work demands and grades (see, e.g., Grigsby 2009; Moffatt 1989; Nathan 2005). Arum and Roksa reported that more than 90 percent of students said they had talked to a professor about grades, but only one-quarter said they had talked to a professor about ideas presented in class. A majority of the 2,400 college students in the Arum and Roksa study said they had not taken a course during the previous term that required a total of twenty pages of written work, and 25 percent said they had not taken a course that required even forty pages of reading per week. Arum and Roksa concluded:

> Given the small amount of time students spend studying, it is no surprise that they are not learning much. This is partly a consequence of lax demands and expectations, but it is careless to think that simply increased faculty demands will produce greater learning in higher education. The college experience is perceived by many students at the core as a social experience. The collegiate culture emphasizes sociability and encourages students to have fun, to do all the things they have not had a chance to do before or may not have a chance to do after they enter "the real world" of the labor market. (2011, 131)

The system of low expectations existed because it served the interests of all major actors

who were in daily contact with the classroom. A large proportion of college students—perhaps a near majority—saw college as a period of fun, friendship, and personal development before they began adult life. They hoped their investments in college-level training would pay off in the labor market, of course, but many assumed, perhaps quite realistically, that credentials themselves would add value, not what they learned in college. While faculty members were interested in making their classes lively and interesting, they also wanted to preserve time for research, correspondence, committee work, and other socioprofessional activities. Challenging requirements and multiple assessments added time to their preparation and created discontent among utilitarian-minded students. Nor were administrators particularly interested in adding more challenging material to the undergraduate curriculum. On the contrary, administrators were usually more interested in reaching enrollment targets and raising retention and graduation rates than in encouraging challenging course work or requiring students to demonstrate cognitive growth (Bok 2006; see also Arum and Roksa 2011, 141).

The climate of low expectations does not accurately describe the practices common at the leading research universities and liberal arts colleges or in some of the more demanding disciplines, such as engineering, math, physics, philosophy, and foreign languages. Instead, it seems to have been particularly characteristic of occupational-professional programs outside of engineering where reading and writing requirements were weakest (Arum and Roksa 2011, 104–9). Yet some of the building blocks of cognitive gain may have been in short supply across the board. In a study of University of California undergraduates in 2008 my research team and I found that *no discipline* stood out in the frequency with which student majors reported analytical and critical thinking experiences in their classes, such as comparing two contrasting perspectives, assembling evidence to support an argument, or breaking down arguments into their component

parts to assess the validity of each one (Brint, Cantwell, and Saxena 2012).

Teaching for Understanding

The forces of the new progressivism commanded impressive organizational tools and a relatively easy-to-implement checklist of reforms to attach to existing curricula. The same could not be said of the much more ambitious, but less completely realized project of the Carnegie Foundation for the Advancement of Teaching under Ernest Boyer's successor, Lee S. Shulman. Under Shulman's leadership, the Carnegie Foundation embarked on a program to redefine and realize Boyer's vision of a scholarship of teaching. These efforts eventually steered the foundation away from the tenets of the new progressivism to a deeper inquiry into the aims and methods of undergraduate teaching. Shulman's approach came to share only part of the faith of the new progressivism in the power of student engagement. Engagement, he wrote, "is not enough." "Understanding is not independent [of engagement] but is an additional standard" (Shulman 2004, 56).

For Shulman, all good teaching was built, in the first instance, on subject matter mastery. Shulman emphasized, in addition, "pedagogical content knowledge"—the special materials and methods tied to knowledge-making in the disciplines, such as work with primary textual materials in history, surveys and ethnography in sociology, and diagnostic clinical rounds in medicine. Based on this knowledge and these disciplinary resources, teaching and learning could be conceived as an interactive process of bringing "something inside" of the teacher out in a methodical and powerful way—and of bringing "something outside" of the student, the lesson, into strong relief in students' consciousness. In all good teaching, methods of expression and bases of apprehension and understanding were consequently closely linked (Hutchings and Shulman 1999).

23

Shulman emphasized that the first obligation of the teacher is to determine what students know and can do, as well as their interests and passions. Working from these bases, Shulman and his associates (Huber and Hutchings 2005) advocated that teachers create "cognitive apprenticeships" in which students were asked to make their mental processes accessible to their fellow students and teachers and to work toward more expert understandings of course materials. Through a process of "uncoverage," teachers were encouraged to focus their first lessons on ideas and concepts that were both difficult to grasp and fundamental to subsequent learning in the class.

Teachers made their own thinking accessible to students by explicating the "intermediate processes" of understanding—the understandings that are employed habitually by expert learners but often hidden in the process of instruction. These could include, for example, explicit discussions of the flow of an argument, the translation of terms no longer in wide use, or a detailed, step-by-step interpretation of the architecture of a statistical table. Other techniques for making knowledge accessible included slowing down students' reading; eliciting students' descriptions of their thinking about passages in the text; administering oral rather than written midterms; employing structured online discussions to create learning communities oriented to key issues and ideas in a course; and posting examples of beginning, intermediate, and advanced understandings of texts with detailed explications of the major differences between these levels of mastery. Similar pedagogies were developed for mathematics—for example, in James Sandefur's "think alouds" in which math students were asked to describe, step by step, how they were thinking about a problem as they worked through its solution.

Shulman argued that students should demonstrate competence by performing skills in front of their teachers and classmates rather than by passively absorbing information. For

Shulman, the pathologies of learning—amnesia (forgetting what was just learned), fantasia (misperceiving the lesson to reinforce existing knowledge), and inertia (inability to use knowledge in new contexts)—were ultimately issues of ownership. Understanding implied ownership and the sense of ownership typically required performing what one knew.

Research on teaching for understanding has expanded beyond the ideas developed and promulgated by Shulman and his colleagues, as I will show in the concluding section of this chapter, but the Carnegie Foundation under Shulman's leadership deserves credit for introducing new insights about teaching for understanding and for popularizing early work in the sciences of learning that provided conceptual and evidentiary bases for these insights. Table 8.1 provides a partial list of elements of teaching for understanding as developed by the Carnegie Foundation and augmented by later researchers.

Table 8.1
Characteristics Associated with Teaching for Rote Memorization
and Teaching for Understanding

Teaching for Rote Memorization	Teaching for Understanding
Teaching is based on textbook coverage	Teaching begins with what students know often based on concept inventories
Instructor dominates class time	Student motivations/interests are assessed
Instructor emphasizes points students need to know for tests	Learning objectives are specified to students
Supplementary materials reinforce key points of lecture	Teaching builds from most fundamental theories/concepts/ideas
Review sheets consist of key points made in class that will appear on tests	Intermediate processes in understanding are discussed in step-by-step fashion
Students drill each on what they think the instructor wants them to know	Instructors discuss the process by which concepts/models/principles developed
Assessments allow students to pass by repeating key points from lecture	Students are asked to give reasons for why they come to conclusions
	"Think alouds" and "difficulty papers" may be used to assess student understanding
	Detailed feedback is given on student work
	Students are encouraged to meet with the instructor to discuss their work

25

Students have the opportunity to redo their
  work based on feedback
Students are asked to perform what they
  know in class presentations, as well as on
  tests
Post-test concept inventories are used
  to assess learning gains

Growing out of the Carnegie program, Scholarship of Teaching and Learning (SoTL) colloquia sprouted on hundreds of college and university campuses, as did a number of impressive websites devoted to this scholarship. The colloquia took up such matters as visually effective presentation of lessons, new ways to assess student learning, uses of technology to improve pedagogy, the impact of learning communities, and other topics consistent with the Carnegie agenda under Shulman's leadership. The SoTL philosophy was not an industrial search for better systems but rather an apprenticeship system for craftsmen, based on sharing the distinctive visions of master teachers.[6]

Shulman's work was widely cited, but the organizational apparatus Carnegie used to spread these ideas showed neither the panache of the AAC&U campaigns nor the reach of NSSE. Instead, an artisanal model, built on networks of sympathetic practitioners, prevailed. This approach generated fresh insights about teaching and learning—insights with the potential to create more effective college teachers. But its insistence on "scaling down" through small-scale actions of unusually committed practitioners was destined to create islands of improved practice in a sea of relative indifference. According to Mary Taylor Huber and Patricia Hutchings,

> The key is not the scale and scope but the care and thoughtfulness of the work, its capacity to change thought and practices, its generosity, even, perhaps, its power to surprise and delight (2005, 30)

Whatever the merits of this argument, it led to relatively thin penetration of CFAT's "pedagogies of understanding."[7] Russell Edgerton, who did so much as a program officer at the

26

Pew Foundation to promote the Carnegie program, concluded that more than two decades of reform activity resulted in "neither professional nor institutional transformation" (Edgerton, personal communication).[8] For all of its shortcomings as a campaign for change, CFAT's ideas about teaching for understanding, and vehicles such as the SoTL colloquia, did help build a cadre of researchers interested in applying the tools of inquiry to the problems of pedagogy

The Outcomes Assessment Movement

In contrast to teaching reform, outcomes assessment can be defined as a response of state legislatures and regional accrediting bodies to the perception that colleges and universities had not done enough to ensure that students were learning course materials and essential academic competencies. Where the teaching reform movement took root in foundation-supported advocacy organizations, the outcomes assessment movement was promoted primarily by the states and the federal government, abetted by the regional accrediting associations.[9]

Fledgling efforts to encourage institutional assessment of learning outcomes began in the 1970s. The Educational Testing Service fielded the first open-response test of core skills, Academic Competencies in General Education, at 140 institutions, but it was later abandoned owing to the tendency of institutions to magnify small pre-/post-test differences and the test's unreliability in the mid-ranges of scoring (Adelman 2007). By the mid-1970s, twenty states had introduced minimal competency testing for graduating seniors, mirroring popular high school exit exams (Gilman 1978). Calls for action continued in the early 1980s, as the presidential report *A Nation at Risk* (1983) documented the shortcomings of U.S. primary and secondary education in the face of increasing competition from East Asia.

Three years later, the National Governors' Association took a stand. *A Time for Results* (1986), a key document of the period, stressed the same fears about the competency of U.S.

27

college graduates and the same looming threat of Asian competition. It noted that U.S. higher education had set a new standard for access but observed that "access without quality is a cruel deception." In the document, a subcommittee of governors, led by John Ashcroft of Missouri, questioned assumptions about higher education: "Learning is assumed to take place as long as students take courses, accumulate [credit] hours and progress satisfactorily toward a degree." But, the subcommittee observed, "tests of elementary and high school teachers show that the BA is not a guarantee of even basic literacy, let alone competence." The report also cited, with little documentation, "substantial levels of dissatisfaction" among employers about the skills of college graduates. The report advocated systematic programs using multiple measures to assess undergraduate student learning, and it cited with approval institutions like Alverno College that had pioneered systematic assessment in the 1970s. It also applauded the Southern Accreditation Commission for being the first of the regional accrediting bodies to require an assessment component for reaccreditation.

Performance Funding: The First Wave

In the 1980s, state financial resources were becoming conditioned upon institutional performance in specified areas. These included student retention and graduation rates, student scores on licensing examinations, job placement rates, faculty research productivity, and measures of undergraduate access and campus diversity (McLendon, Hearn, and Deaton 2006). Between 1979 and 2007, twenty-five states enacted performance funding (though ten of those states dropped it over the years) (Burke and Minassians 2003; Dougherty and Reid 2007). Performance funding proved costly to implement, susceptible to institutional manipulation of performance measures, and subject to reversal under new administrations or when unstable state finances caused deep cuts in regular higher education funding (Burke and Serban 1998; Dougherty and

Natow 2009; Shulock and Moore 2002; Zumeta 2001).

Nevertheless, new demands for accountability, including direct assessment of student learning, gradually gained ground during this period. A 1987 report of the Education Commission of the States showed that two-thirds of states had initiated some form of required student assessment. However, many states used minimal competency measures at graduation, or even more indirect measures, such as graduation rates and pass rates on professional licensing examinations. Although assessment of student learning was in the air, few knew how to test directly for student learning outcomes in a cost-effective, relatively unobtrusive way. The large testing companies, ACT and ETS, thought they did know how to do it. They geared up for the new era by introducing or revamping multiple-choice tests, the Collegiate Assessment of Academic Proficiency (CAAP) and the Measure of Academic Proficiency and Progress (MAPP), respectively, that institutions could administer to their freshmen and seniors to determine the institution's "value added" to student academic competencies.

The "Learning Paradigm"

To the extent that a manifesto existed for the outcomes assessment movement, it was produced by two state college professors in California, Robert Barr and John Tagg. In a widely cited article from *Change* magazine, Barr and Tagg sought to shift thinking in academe from an "instruction paradigm" to a "learning paradigm":

> The paradigm that has governed our colleges is this: A college is an institution that exists to provide instruction. Subtly but profoundly we are shifting to a new paradigm: A college is an institution that exists to produce learning. This shift changes everything. (1995, 13)

The idea of a shift to a learning paradigm resonated strongly among state educational bureaucrats and in the world of higher education policy analysts.[10]

29

Over the next five years, a chorus of influential voices called for measurement of student learning outcomes and influential practitioners created demonstration projects to show how this measurement could be done. In 2000, the National Center for Public Policy and Higher Education, funded by several major foundations and led by the former governor of North Carolina and educational reformer James B. Hunt, began to publish report cards about state higher education performance, including "incomplete" grades for all states on student learning. In the same year, ABET, the accrediting agency for engineering schools, began its Engineering Criteria 2000 policy requiring outcomes measures and plans for continuous improvement based on results of outcomes assessments. In 2002, the Pew Trusts provided funding to two leaders of the assessment movement, Margaret Miller and Peter Ewell, to demonstrate the possibility of measuring college learning in six states for future incorporation into the National Center for Public Policy and Higher Education's "Measuring Up" reports. In 2003, the Carnegie Corporation of New York and the Teagle Foundation sponsored the development of a new type of test of core academic skills, the Collegiate Learning Assessment, based on the use of document libraries to solve "real-world" problems. In the same year, the national council of regional and disciplinary accrediting agencies, the Council for Higher Education Accreditation (CHEA), announced a policy of "mutual responsibility" between institutions and regional accrediting agencies for demonstrating student learning outcomes.

An opinion survey published by the Educational Testing Service (ETS) in 2003 discovered evidence of public concerns about educational quality, stronger among political conservatives and high school educated people than among liberals and those with college degrees. Primed by questions linking costs to quality assurance, a majority surveyed by ETS agreed that colleges should provide evidence that they were producing the learning results they

promised, if they were going to continue to raise costs (ETS 2003). In 2004, the Business-Higher Education Forum argued for the first time in favor of assessments of student learning outcomes. Also in 2004, Miller and Ewell published their six-state report showing that states could demonstrate student learning outcomes through a variety of methods, including proficiency benchmarks. In 2004, the State Higher Education Executive Officers (SHEEO) launched a National Commission on Accountability in Higher Education, chaired by former secretary of education Richard Riley and former Oklahoma governor Frank Keating, both Republicans. The report they produced in 2005 concluded that most state systems "do not meet their intended purpose to improve and to provide evidence of student learning" and endorsed collection of data on student learning outcomes (National Commission on Accountability in Higher Education 2005).

The Accrediting Bodies Respond

By the mid-1990s all six of the regional accrediting agencies had policies in place requiring institutions to demonstrate not only that they were tracking conventional measures of student success, such as four- and six-year graduation rates, but also that they had mechanisms in place to achieve established goals for student learning. In 1998, Congress formalized this commitment by making student achievement the first of nine areas in which the regional accrediting agencies were required to have standards.

While following federal directives for recognition, regional accrediting agencies have buffered institutions from state pressures for standardized testing. Some allowed institutions to take responsibility for assessing and achieving a unique set of learning outcomes that they establish for themselves. Others identified a core set of learning outcomes that ought to be examined by all institutions. These typically encompassed, at a minimum, critical and analytical

thinking, written expression, and quantitative reasoning. Institutions and departments were granted considerable autonomy so long as they provide evidence that they are establishing learning objectives and developing ways to assess and report the achievement of these objectives. This permitted a variety of assessment approaches, ranging from the presentation of portfolios of student work to requirements for integrative research papers in senior capstone courses. Others built in learning objectives to required courses and required samples of work from these courses or adopted exit examinations as a way of determining whether learning objectives have been met.

Although the regional accrediting bodies developed elaborate procedures to ensure that institutions did more than pay lip service to their demands for evidence of student learning, accrediting requirements were nevertheless often treated by faculty members as an encumbrance requiring the appearance of compliance without deeper commitments to the goals of evaluating student learning in a more rigorous or consistent way at the programmatic rather than the course level. The limited resources and experience of accrediting agencies also encouraged high levels of institutional latitude; most, if not all, of the regional accreditors lacked experience in evaluating evidence of student learning or the qualifications to establish clear standards by which to do so (Ewell 2001a).

Even so, by fostering a common demand for evidence about student learning, the regionals created much more attention to student learning outcomes than had existed before. In 2009, the National Institute for Learning Outcomes Assessment (NILOA), housed at the University of Illinois, fielded a study of the incorporation of assessment instruments. The study was funded by the Carnegie Corporation, the Lumina Foundation, and the Teagle Foundation. Officials at half of U.S. two- and four-year institutions responded to the survey, and the vast

majority (92 percent) said they were engaged in institution-level assessments of student learning. Most said they were using survey instruments like NSSE, but 39 percent said they were also using standardized tests of general knowledge and skill like CLA. At the program level, four of five respondents said they were assessing student learning outcomes in at least one program, and here portfolios of student work dominated. Most said that accreditation was the primary driver of their interest in assessment (Kuh and Ikenberry 2009).

The Spellings Commission Report

In 2004, George W. Bush's secretary of education Margaret Spellings appointed a Commission on the Future of Higher Education, chaired by Texas businessman Charles Miller, to recommend reforms in higher education accountability. In 2006, the commission issued its final report, *A Test of Leadership*, which was highly critical of the performance of America's colleges and universities. The report dismissed previous efforts to bring accountability for student learning outcomes.

> Despite increased attention to student learning results by colleges and universities and accreditation agencies, parents and students have no solid evidence, comparable across institutions, of how much students learn in colleges or whether they learn more at one college than another. Similarly, policymakers need more comprehensive data to help them decide whether the national investment in higher education is paying off and how tax payer dollars could be used more effectively. (Spellings Commission 2006, 14)

The commission advocated measuring student achievement on a value-added basis that took into account students' previous achievements when assessing outcomes. It stated that this evidence should be made available to consumers and policymakers in an accessible, understandable way, and it proposed that "meaningful" interstate comparison of student learning be encouraged and implemented in all states (ibid., 4).

The specter of high-stakes testing haunted many in academe, who argued that such tests

would yield little of value for students studying such a wide variety of disciplines (see, e.g., Chatman 2007; Hawthorne 2008). The only way to test learning would be discipline by discipline, these educators argued, and this seemed an impossible task given the limited resources of colleges and universities and the limited capacity of state educational bureaucrats to grade such a wide variety of tests.[11] Leaders of the testing movement countered that tests of general skills were an important, if not the only important, measure of student achievement in college. Instead of relying on one test, they argued, multiple forms of assessment would be necessary—some to assess general skills, others to assess disciplinary knowledge, and still others to assess the "soft skills" required in leadership positions (see, e.g., Ewell 2004; Shulenberger 2008).

The Bush administration proposed that the federal government take a larger role in quality assurance. After extensive lobbying by the higher education associations, Senator Lamar Alexander, the chair of the committee responsible for reauthorization of the Higher Education Act, was convinced to allow the existing system of regional and professional accreditation to continue and to bar the federal government from prescribing standards that these agencies were required to use in assessing institutional effectiveness. But, in exchange for his support, Alexander insisted that higher education institutions themselves take on the responsibility to measure student learning outcomes in a serious way. Alexander's intervention led to the creation of the Voluntary System of Accountability (VSA), organized, with support from the Lumina Foundation, by two of the leading higher education associations. VSA set as an explicit goal the development of a system of accountability that would "facilitate comparisons of learning outcomes among institutions of higher education" (Millett et al. 2007, 2).

The VSA ultimately failed as an accountability mechanism. Of the more than 300

34

institutions participating in VSA as of fall 2009, less than one-third reported results of "core academic skills" using one of the three authorized testing instruments. Of the reporting institutions, the expected two-thirds reported results within a standard deviation of the mean for institutions with similar student academic ability profiles, but among the remaining institutions three times as many reported results "above" (one standard deviation above the mean) or "well above" (two standard deviations above) expected as those reporting results "below" or "well below" expected. Indeed, only five of 104 reporting institutions said that they were performing below expected levels, a statistical impossibility.

The Limited Outcomes of Outcomes Assessment

The states and the regional accreditors proved to be strong advocates of assessing student learning outcomes but weak implementers. By the end of the period, the states had been persuaded to defer to the regional and professional accrediting associations to provide quality assurance and to the VSA to experiment with the construct validity of several tests of general intellectual skills and to use these tests to monitor the "value added" of institutions.

Neither the regional accrediting bodies nor the VSA transformed the college classroom by demanding evidence of student learning outcomes. They encouraged richer discussions about learning objectives, but the regional accrediting agencies, for the most part, allowed institutions and departments to formulate their own objectives and to choose their own methods for demonstrating results. These requirements did not change practice as much as reformers hoped, because they allowed departments to be their own prosecutors, judges, and juries. Similarly, the learning outcomes component of VSA continued to be slow to get off the ground. Thus, while national and trans-institutional actors succeeded in shaping the environment of discourse, their efforts met both passive and active resistance whenever they have attempted to prescribe tough

35

standards for the assessment of student learning outcomes.

The sociologist Jal Mehta (2007) offered a plausible explanation for the fierce rhetoric but limited follow-through of the states in assessing student learning outcomes. In Mehta's view, higher education has been protected from accountability pressures by its reputation for quality and expertise, and by its larger private sector, which is practically immune from state accountability pressures. One might add political factors to this explanation. These political factors included the ability of higher education advocates to exploit doubts about the effectiveness of K–12 reform, as represented by the No Child Left Behind Act of 2000, partisan turnover in the governing coalitions of the states, and, in particular, the capacity of the higher education associations and regional accrediting bodies to assure key legislators that they would implement accountability measures responsive to public interest in quality assurance. Most state governments were, in the end, willing to accept these assurances, in part because they did not want to spend scarce state resources on enforcement mechanisms. The stakes were just not high enough to take policing out of the hands of the regional accreditors.

Performance Funding: The Second Wave

Performance funding, however, continued to interest state policymakers, even after scholars and policy analysts identified persistent problems with its implementation and outcomes, including performance-based allocations that were too low to stimulate change and one-size-fits-all metrics that were not tailored to the circumstances of different types of colleges and universities. These findings led to changes in performance-funding formulas, with states raising the levels of funding based on performance and allowing institutions at different selectivity levels to adjust metrics to suit their student base (Li 2014). In spite of its rocky history, performance funding consequently gained momentum in the 2000s, winning the backing of the Gates and Lumina

36

foundations and their networks. By 2014, twenty-five states either used performance funding or were planning to incorporate it. Tennessee continued to be its most persistent and aggressive advocate, allocating 100 percent of higher education funding based on performance indicators.

Even so, the expected outcomes failed to materialize. In a review of the extensive state-level literature on outcomes, the higher education policy analysts Kevin Dougherty and Vikash Reddy found "no statistically positive impacts of performance funding on six-year graduation rates in public four-year colleges" (2011, 27). In her examination of 467 institutions between 1997 and 2007, Jeong Cheol Shin (2009) reported similar results: no significant change after the introduction of performance funding. David Tandberg and Nicholas Hillman (2014) were among those who reported the most positive results; using sophisticated statistical techniques on state-level data from the period 1990 to 2010 they found small but statistically significant improvements in graduation rates in the seventh and eighth years after performance-funding implementation. Whether political winds would allow performance-funding stability up to eight years for these kinds of modest results remained an open question. And the unintended consequences of doing so also remained an open question; reports continued to trickle in of institutions adjusting graduation requirements downward in efforts to improve their performance profiles (Li 2014).

A Breakthrough Science of Learning?

Thus the most obvious consequences of two decades of reform were the diffusion of active learning pedagogies and surface-level adoption of relatively weak accountability measures to assess student learning outcomes. These limited achievements were clearly not enough to change the social relations of learning prevailing in most college classrooms.

But one development during the period did contain the potential to do so. The sciences of

learning gained momentum in the 1990s and put new life into academe's self-examination of teaching practices. Little by little, change in teaching practices began to occur based on the creation and diffusion of new knowledge about effective teaching practices. Natural science educators, particularly physicists, took the lead in these efforts. Thanks to what now amount to thousands of research studies relating student learning to teaching practices, the techniques associated with learning gains in undergraduate science and mathematics courses are well known. Many of these practices have proven to be transferrable to courses outside the natural sciences, albeit not as often to the arts and humanities where pedagogies based on discussion and interpretation loom larger.

Slightly different emphases have emerged among the various groups that have sought to reform practice based on the results of research, but some broad commonalities are evident among them. The alignment of course content and course assignments with stated learning objectives is one such commonality. The use of active learning techniques is a second. The creation of high-energy and inclusive classroom environments is a third. The use of materials that require students to engage in analytical and critical thinking rather than rote memorization is a fourth. And the use of frequent assessments targeted to course learning objectives is a fifth. (For overviews, see Ambrose et al. 2010; Bransford, Brown, and Cocking 2000; Freeman et al. 2014; Froyd 2008; Wieman 2012.)

Teaching improvement became a policy priority by the end of the period, as indicated by the advocacy of the American Association for the Advancement of Science, the National Academies, the President's Council of Advisors on Science and Technology, and virtually all of the major higher education associations for the science of learning. Networks of expert practitioners sponsored by these organizations (and by some individual campuses) fanned out

38

across academe to show how evidence-based teaching practices could be implemented to increase students' learning gains. Through these means, the principles developed by cognitive researchers were embedded in course redesign programs and in instructor certification programs such as the one cosponsored by the American Council on Education (ACE 2016), as well as in discipline specific efforts such as PULSE (Partnership for Undergraduate Life Science Education) with its heady roster of medical and federal sponsors (PULSE 2016).

In the following sections I will discuss the tools developed by learning scientists: those that are intended to increase student participation in learning, to improve students' accountability, and to help instructors to teach more effectively for understanding rather than rote memorization.

Participation Tools

The research literature showed that one strand of the new progressivism was entirely correct: students tended to learn more when they were actively engaged through classroom participation in the topics under discussion. Lecture halls had proven to be the bane of undergraduate education because of their tendency to bring out the worst aspects of mass higher education: the complete domination of classroom time and focus by instructors and the passivity of many students in the face of this control.

In a now classic article, the physicist Richard R. Hake (1998) compared traditional lecture-based classrooms in introductory mechanics courses to those based on what he called interactive engagement. Hake used two well-validated tests, the Halloun-Hestenes Mechanics Diagnostics test and the Force Concept Inventory. Instructors gave these tests to students prior to the beginning of class and then again at the end of class. Hake used changes in scores on the tests (pre- and post-test) to determine average gains for courses taught in the traditional lecture mode

and those taught using interactive engagement techniques. The traditional classrooms were based on lectures typically followed by a short time for student questions. The interactive-engagement classrooms, by contrast, were based on mini-lectures or demonstrations followed by breakout sessions in which groups of students discussed a problem posed by the professor related to the topic under discussion. Some groups were then asked to report out on their conclusions. The professors corrected groups that had come up with wrong conclusions and elaborated on other groups' correct conclusions. Hake compared pre- and post-test results on the Force Concept Inventory for more than 6,500 students in 67 introductory physics classes in 14 high schools, 16 colleges, and 32 universities. Average scores not surprisingly varied considerably from campus to campus, but the key comparative results were consistent and cumulatively persuasive; students in the interactive-engagement classes outperformed students in the traditional classrooms, with average gains almost two standard deviations above those achieved by students in traditional classrooms.[12] The time allowed for students to interact with one another around a problem and the professors' responses to their solutions evidently helped cement the learning of key concepts and relationships beyond what could be expected in a traditional lecture format. Concept inventories were subsequently developed for numerous fields, including astronomy, chemistry, engineering specialties, evolutionary science, geosciences, and statistics and applied in studies similar to the one reported by Hake (see, e.g., Smith, Wood, and Knight 2008). The results of studies using these concept inventories have confirmed the learning gains associated with interactive engagement as opposed to traditional lecturing.

The consistency of the advantage for interactive engagement helped lead a transformation in physics teaching and subsequently in many of the natural sciences and engineering. The physicist Eric Mazur (1997) was another leader in what he called the "peer instruction"

40

movement. Mazur recounted that his sponsorship of new approaches to classroom instruction came from his realization that his introductory Harvard students were performing well on examinations because they could apply memorized equations appropriately to problems, but most did not have the foggiest idea about the underlying concepts and principles that made the equations work. They were dutifully working through difficult assignments without truly understanding what they were doing or why they were doing it.

During the period researchers focusing on interactive engagement developed a variety of techniques to foster student participation in class. The techniques built on and in some cases reconfigured the small-group discussions on which Hake and Mazur based their studies. One closely connected variation was the "think-pair-share" technique, which is based on teaming up two students to compare their ideas about a question posed by the professor and then having one of the students report out if the group is called upon. (Those who have used this technique emphasize the importance of allowing some silent time for more introverted students to compose their thoughts prior to the discussion with their partners.) Instructors also learned to begin their classes with a "hook" that produced discussion among the students. Polling for student responses to questions, either using electronic means or simply raised hands, allowed professors to question students on the reasons for their diverging responses.

Some professors with classes of students who were reluctant to answer questions walked the aisles of their lecture halls and offered extra points to students from different quadrants of the room who were willing to venture an answer to a question. Others used the jigsaw technique in which each member of a group was tasked with learning one segment of a multipart assignment and segment experts from the several groups consulted one another prior to the presentation of the segment to other members of each of their jigsaw groups. Another colorfully named

technique, the fishbowl, placed a group of students in front of the class to present a discussion of an assignment and permitted those outside the fishbowl to join those in it by asking questions or in some cases physically joining. Many more familiar approaches to interactive engagement also grew in popularity during the period, including debates, oral presentations, and contests for the best illustration of important concepts, ideas, or principles (see DeAngelo et al. 2007) (see table 8.2).

Table 8.2
Teaching Practices Associated with Student Passivity
and Student Engagement

<u>**Student Passivity**</u>	<u>**Student Engagement**</u>
More than 80 percent of class is lecture	Lecturing constitutes less than 60 percent of class time
Instructor stands at podium or in front of class	Instructor moves around the lecture hall asking questions
Instructor makes few efforts to engage students	Mini-lectures are followed by small group break-outs
Instructor plans few activities other than lecture	Instructor attempts to bring in all students by offering extra points for answers by students in all quadrants of the class
Videos/Films are main supplementary materials	
Students are not called upon to answer instructor's questions	Other student discussion techniques are employed (e.g. think-pair-share, jigsaw, fishbowl)
Question time is limited to a few minutes at the end of class	Demonstrations, simulations and videos are used; students are first asked to predict outcome
Slides on topics covered are available prior to class	Class includes opportunities to conduct hands-on research and/or to create a creative work and report on results
	Competitions may be employed (e.g. debates, best illustrations of concept, best posters)

Studies I conducted of University of California students indicated that first-generation

college students, particularly Asian American and Hispanic students, were less inclined than others to say that they participated in class (Brint, Cantwell, and Saxena 2012; Brint and Cantwell 2014). The sciences of learning underscore how important it is for instructors to encourage students to make the effort to participate actively in class, even in the face of their inhibitions. Increased learning has been the primary motivation to adopt peer instruction techniques, but greater equity in learning may be an important byproduct.

Accountability Tools

In the consumerist culture of the period, students complained regularly that too much was being required of them, but the truth seems to be that too little was being required of them. The time-on-task literature helps to show why. Time on task has been associated with stronger performance on tests in dozens of studies of K–12 classrooms. For similar reasons longer reading and writing assignments are associated with larger gains in analytical and critical thinking among otherwise similar students in higher education. The effort required to make sense of difficult prose, to wrestle with ideas, and to compose a longer paper can build cognitive capacity in much the same way that challenging workout regimens build the body's fitness.

Of course, students are regularly assessed on their class performance through paper assignments, quizzes, and examinations. Even so, the repertoire of assessment did not seem to be fully adequate to ensure student accountability for learning. Many students did not read for class and were unprepared to discuss reading materials. For this reason, college instructors turned to online reading quizzes, in-class reading quizzes, or brief responses to prompts about the assigned reading. One notable experimental study showed that daily online reading quizzes significantly improved student performance on final exams in an introductory psychology course while at the same time reducing achievement gaps between students from high- and low-income backgrounds

43

(Pennebaker, Gosling, and Ferrell 2013). Dozens of similar findings have been reported elsewhere in the literature (see, e.g., Bell 1997; Johnson and Kiviniemi 2006; Marcell 2008; Padilla-Walker 2006).[13]

The disallowing of laptops and other electronic devices in the classroom has proven a more controversial approach to improving student accountability. Some students used their devices to take notes and to look up materials discussed in class online, but many others sat in the back rows of lecture halls and used them instead to post on social media, to catch up with the sports news, or to shop. Disallowing electronic devices made an emphatic statement about the priorities of the instructor, but it frequently also led to resentment. As an alternative, instructors ensured that electronic devices were being used for class-related purposes simply by walking the aisles while lecturing and docking points for those who were using their devices for non-class-related purposes. The elimination of screen-based distractions helped re-create the classroom as a "sacred space" in the Durkheimian sense, one that unlike normal life requires ritual respect for the "totemic" object of the subject under discussion and the focused attention of participants (see table 8.3).

_____

Table 8.3
Teaching Techniques Associated with Low and High
Student Accountability for Learning

Low Accountability	High Accountability
No checks on whether students have done reading prior to class	Quizzes or prompts on reading are given prior to class
Attendance is not taken	Attendance is taken and points are reduced for non-attendance
Students are allowed to come in late and leave early without instructor's permission	Students are not allowed to come in late or leave early without instructor's permission
Laptops and devices are not monitored for class related purposes	Laptops and devices are not permitted or are monitored for class related purposes
Slides for class are available prior to	

class meeting	Many assessments are given
Assignments are not challenging	Assignments include relatively
Few assessments are given	complex reading and lengthy papers
Feedback on assignments and exams is very limited	Feedback on assignments and exams is extensive
Students are not required to attend office hours to review work if grades are low	Students are required to attend office hours if grades are low

The science of learning showed that student engagement and student accountability were both essential supports for the goals of teaching for understanding and cognitive development. The cultures of the disciplines, however, were divided so that instructors in more interpretive fields tended to emphasize high levels of classroom participation as the primary indicator of student commitment to learning and those in more quantitative fields tended to emphasize dutiful adherence to difficult work as the primary indicator of student commitment. With my colleagues Allison Cantwell and Robert Hanneman, I characterized these as the "two cultures" of student academic engagement (Brint, Cantwell, and Hanneman 2008), mirroring at the classroom level C. P. Snow's (1959) famous distinction between the two cultures of the humanities and the sciences. During the period, the two cultures slowly began to merge into one, as humanists looked for greater accountability and scientists began to incorporate higher levels of student participation.

Tools for Understanding and Problem Solving

Researchers in the science of learning built on many of the principles discovered by cognitive scientists and promoted by Lee Shulman and his colleagues at the Carnegie Foundation as "teaching for understanding." They demonstrated that effective learning objectives, such as showing the connection between course materials and students' career goals, can reinforce students' willingness to commit to cognitive mastery. They also emphasized the importance of

showing students the process by which concepts, models, and principles were developed and something also about the personalities and conditions under which they were developed. They continued to emphasize the value of uncovering every step in students' thinking as they worked through problems. Researchers emphasized the importance of asking students why they came to particular conclusions and the necessity for quickly and considerately correcting those reasons when they were in error.

Learning scientists also emphasized the employment of exercises that can help students improve their analytical and critical thinking skills. These techniques have in fact been well known for decades, but many professors stopped using them in the face of increased student demand for easy-to-digest-and-remember content. They include assignments that require students to compare and contrast two or more perspectives on a topic of interest. To approach an assignment like this, students must first show that they understand the fundamentals of the analytical models under consideration. They must show that they can research the outcomes of cases in which each of the models was applied, and they must show that they can draw at least tentative conclusions from what their research has revealed. In this way, four essential elements of thinking come into play: definition of terms, evidence collection, application of concepts to new situations, and drawing conclusions based on evidence and analysis. The practice of preparing, performing, and analyzing experiments in science labs derives from the same tradition of deep thinking exercises, and it engages a similar set of mental operations: understanding the problem (definition), researching (through experimentation), analyzing results, and coming to conclusions based on results.

Researchers went beyond the precepts of teaching for understanding to consider also the teaching of problem solving skills. Problem-based learning also has a long history (see, e.g.,

Bruner 1961), but it gained a new impetus from the idea that students should show what they have learned in "real-world" situations. "Real-world" problem-solving exercises were intended to explicitly tie classroom learning to workplace dynamics. Many problems in workplaces resemble problem-based learning activities and call for the same work group engagement with problem definition, research on previous approaches, designs for interventions, and evaluation of the consequences of the enacted solutions. Problem-based learning approaches consequently focused on ambiguous problems that required students to attempt a variety of solutions, often using research to guide problem-solving, and sometimes also to design methods to determine whether the solution worked. An architecture class might, for example, work on building designs that maximize opportunities for inhabitants' interactions or an urban studies class might work on the best ways to determine the health care needs of a poor community and cost-efficient ways to provide for those needs once assessed. Well-designed problems of this type created a shared culture of learning and collaboration while at the same time engaging students' critical and analytical thinking skills.

Students exposed to problem-based learning exercises were required to go beyond the dutiful but surface work of repeating what teachers wanted them to know to the more challenging but also potentially more empowering work of applying what they had been learning to new problems that did not have immediately apparent solutions. These approaches required students to demonstrate persistence because they did not open themselves to readymade answers. As Shulman had argued, presentations of the work accomplished helped cement learning while providing students with the experience of receiving feedback on the approaches they had taken and the results of those approaches. In the professions, traditions such as medical rounds and moot court require public performance of knowledge because this performance is what counts in

practice. Public performance can be a means of cementing knowledge ownership in the arts and sciences as well (Shulman 1997). It is one reason why posters and oral presentations became staples of the hundreds of undergraduate research symposia held annually on college and university campuses and why they have been adopted by some classroom instructors as well.

Given the new emphasis on problem solving, it is not surprising that the trend toward student participation in undergraduate research advanced during the period. On many research university campuses, more than 50 percent of students engaged in an undergraduate research project under faculty and/or graduate student supervision at some point during their college careers either for credit, for pay, or as volunteers. Well-supervised undergraduate research could be a win-win for students and faculty. For students, it fostered deep engagement in problem solving; for faculty members, it provided the opportunity to assist students by using the skills and passion that they brought to their own research projects. Dozens of research universities held spring undergraduate research symposia, with participation reaching 1200 at the University of Washington, Seattle by the end of the period (University of Washington-Seattle 2018) and very high proportions of graduating seniors participating at institutions such as the U.S. Military Academy (Keith, personal communication).

Indeed, many campuses adopted undergraduate research or creative activity experiences very early in the college career in order to capture student interest in the excitement of discovery and to make the drier learning in the classroom more interesting for its applicability. Advocates of early research experiences hoped that as students saw the payoff of these experiences, they would be more willing to spend time learning the concepts, principles, and tools that professional researchers brought to their work. The leaders in this area began to transform undergraduate education from a "sit-and-listen" experience into a "go-out-and-discover" experience. One of

these leaders, National Academy member Sue Wessler, merged bench and computational science with problem-based mini-lectures in a seamless whole for dozens of students in her first-year "Dynamic Genome" course (Warren 2018). In perhaps the most ambitious effort, the University of Texas-Austin started a program that involved more than 700 first-year science students in faculty-supervised research (University of Texas-Austin 2018),

Reflecting Mirrors for Instructors and Administrators

Since they were introduced in the 1960s, student evaluations of teaching have been used by instructors and universities as the primary means for providing evidence of teaching effectiveness. These evaluations provide valuable evidence only on the elements of teaching that students are in a position to evaluate accurately. Those attributes include whether professors are on time, organized, clear, enthusiastic, and approachable. The biases of student evaluations of teaching are well known: less motivated students give lower evaluations; students sometimes focus on irrelevant features of teaching, such as an instructor's gender or appearance; and students are poor judges of whether they have learned much and whether teachers have been effective in producing deeper levels of understanding. One recent study, based on sophisticated statistical methods, concluded that "student evaluations of teaching (mostly) do not measure teaching effectiveness" (Boring, Ottoboni, and Stark 2016). This study joined a long line of research that came to similar conclusions. A good question is how this cockeyed system ever became institutionalized. Why would anyone expect students to have evidence-based knowledge of the techniques that can help them learn? It would be like asking hospitals to rate the effectiveness of doctors solely on the basis of what patients said about their bedside manner.

The Wieman-Gilbert (2014) Teaching Practices Inventory (TPI) emerged during the period as a promising approach to supplementing student evaluations with evidence-based

49

knowledge about effective teaching.[14] The TPI was developed for science and mathematics courses, but it was used with minor modifications in social science and social science–related courses.[15] Following the completion of the course, instructors simply self-rated their use of research-based teaching practices using a point system developed by Wieman and Gilbert based on their assessment of the relative importance of the practices included in the inventory.[16] The inventory, which took ten to fifteen minutes to complete, included seventy-two items for self-evaluation.

The inventory included many items related to teaching for understanding: It included points for the specification of learning objectives, including for lists of general competencies and for topic-specific competencies instructors expected their students to gain from the class. It also included points for the specification of instructors' "affective goals" for the course, as well as points for providing supplementary materials on websites or in handouts. It awarded points for including discussion of the process by which a concept, model, or principle was developed. It also allocated points for questions that required students to explain their reasoning for coming to a conclusion. It awarded points for assignments in which feedback was given and in which students were allowed to redo their work to improve their grades. It also included points for providing answer keys to students following grading of assignments and for encouraging students to meet with their instructor to discuss questions they answered incorrectly. It awarded points for the use of pre- and post-test concept inventories in which gain scores could be calculated for the class.

The TPI also included points for interactive engagement activities, including the number of small-group discussions during the term and the number of demonstrations, simulations, or videos where students were first asked to predict an outcome. It added points for the number of

questions asked by the professor that were followed by student-to-student discussion. It also allocated points for lower fractions of the class given over to lecturing. Indeed, lecturing had to account for *less than 60 percent* of total class time during the term for instructors to gain a point on the TPI for good practice.

Recognizing the importance of student accountability, the inventory also included points for problem sets and assigned homework that contributed to the course grade. It included points for quizzes or prompts on the reading given prior to the class meeting to ensure that students were prepared for class. It included points for the assignment of papers or projects that gave students a degree of choice over topics. It also included points for the number of assessments used above a minimum of two, based on the principle that students learn more from multiple assessments of their performance.[17]

Resistance to Reform

Faculty members who have not yet embraced these advances in teaching practice continue to express skepticism about the value of following *any* prescription for effective teaching. Part of this skepticism stems from traditions of academic freedom, which legislate against any interference in the classroom. Academic freedom protects the right of professors to express views that may flout conventional wisdom. On these grounds, some argue that expecting professors to use particular instructional methods impinges on academic freedom. And, of course, it is also true that different practices may be appropriate for different disciplines or different classes. Nevertheless, a distinction can be made between content and form. Academic freedom guarantees that the content of a course be left to the professor's expert judgment. But it is reasonable to expect that the forms of instruction be governed by professional standards based on well-established research findings.

Some have argued that efforts to pigeonhole faculty members into a particular teaching style robs them of their individuality and, further, that it is precisely the full expression of the individual personality that most affects students. For those who have argued this position, the conditions for human transformation are created by the match between the expression of the personality of an instructor and a student receptive to that particular personality. What needs to be weighed against this insight is the comfort that such a wide berth offers to instructors who are transforming no one, even as they dominate the great majority of classroom minutes. Teachers who are at the high end of the charisma scale should be left alone to work their magic. However, for professors who are not at the high end of the charisma scale, it is reasonable to expect that evidence-based practices will be given a tryout.

Another argument against the professionalization of teaching is resource driven; skeptics argue that instructors face severe resource and time shortages and any practices that add time to teaching should be rejected on pragmatic grounds. These objections are weighty when they come from instructors who have high teaching loads, large classes, and little institutional support. However, most of the teaching practices in the Wieman-Gilbert TPI require no more than a degree of forethought to put into effect. They are not onerous to adopt, particularly if universities were willing to reward those who made the effort. How difficult, for example, would it be for instructors to spell out learning objectives or to carefully align their course assignments with these objectives?

University administrators can use the same reflecting mirror discipline to determine whether the facilities they provide to instructors, the class sizes they sanction, the incentives they employ to improve teaching, and the metrics they use to measure effectiveness are better aligned with student learning or institutional interests in net tuition. Many institutions have found ways

to enhance learning without imperiling the bottom line. I have mentioned a number of them in this chapter. These institutions can provide practical guidance on the way forward for administrators who are focused on monitoring body flow without sufficient concern for the cognitive development of the students flowing by.

The Medical Comparison

A comparison between education and medicine provides a sense of the distance to be traveled by college and university instructors. If doctors failed to use standard practices in the diagnosis and treatment of their patients, they would receive complaints from their patients and colleagues and could be sanctioned by their employing organizations. By contrast, students do not often complain about teachers who fail to use the teaching analog to standard medical treatment practices, provided they receive acceptable grades. And universities do not sanction instructors for their failure to employ them. But universities could easily decide to reward those who employ well-validated teaching practices with higher marks on the teaching component of their reviews for salary increases and promotions. That would begin to make college teaching a professionalized activity, something that it cannot currently claim to be.

Notes

1 Later administrations of the National Assessment of Adult Literacy to samples made up exclusively of recent college graduates showed no declines in literacy. The National Research Council concluded that the test as constructed could not detect who was proficient in literacy skills (NRC 2005), but one wonders whether recent college graduates are the right sample on which to base this judgment.

2 It can be argued that the causes of this renewed focus on the classroom were similar to those that provoked rethinking of classroom teaching in secondary schools at the turn of the twentieth century: the construction of a mass system, fueled by the incorporation of working-class and immigrant students, in which a majority of students had limited preparation for learning and in which chronically underfunded institutions had limited resources to create powerful learning communities. Expansion, combined with continuing fiscal pressures in the

public sector, encouraged concerns about the effectiveness of college teaching, while diversification led to concerns about the possibility of unequal results for women, minorities, and immigrants.

3 As early as 1969, it had issued a statement crediting minorities for "giving a fresh and compelling impetus to the movement for restoring relevance to academic programs" (AAC 1969.). Its studies on the "chilly climate" for women in college classrooms received national attention in the 1970s and 1980s (see, e.g., Hall and Sandler 1982).

4 NSSE included student self-reports of learning gains in several skill areas. Self-reports showed modest correlations with objective tests of learning gains and cannot be taken at face value as evidence of student learning (see, e.g., Bowman and Seifert 2011).

5 A similar study with more elaborate controls on students' prior achievements also yielded modest or insignificant relationships between NSSE benchmarks and cognitive growth on the Collegiate Assessment of Academic Proficiency (Pascarella, Seifert, and Blaich 2009).

6 Other Shulman-inspired projects led to the creation of websites intended to spread pedagogical practices consistent with the "teaching for understanding" approach. Georgetown professor Randy Bass's *Visible Knowledge Project* website was the most important for advancing and codifying ideas about pedagogies of understanding. His website spotlighted techniques for slowing down and deepening knowledge transmission, for building on core ideas and concepts, and for making teachers' intermediate processes and performance standards visible to students, while revealing students' prior understandings and making their difficulties in understanding course materials visible to teachers.

7 The institution of the Carnegie Academy for the Scholarship of Teaching and Learning (CASTL) was the first of Shulman's organizational vehicles. Pew Foundation funds provided support for a summer academy located at the foundation where successful applicants, approximately fifteen a summer, met together to discuss and develop the ideas from their proposals for improvement of college teaching and learning. The total number of CASTL scholars topped out at fewer than one hundred. By contrast, SoTL colloquia emerged on campuses throughout the country, but they attracted only a minority of motivated teachers to their events. Even at such a highly engaged campus as Indiana University, only about one-quarter of tenured and tenure-track faculty had participated in a SoTL event by 2002, and fewer than sixty people attended these events, on average, on a campus of more than 2,000 faculty members. Some other Carnegie projects wound down by the end of Shulman's tenure as president. The Visible Knowledge Project ran out of funds in 2005 after a decade of pioneering work. Peer Review of Teaching remained operational but attracted a dwindling number of new portfolios after Pew funding ended. Carnegie's Knowledge Media Lab closed its electronic doors in September 2009, though its course portfolio software remained retrievable.

8 The objectives of the Carnegie Foundation changed dramatically with the selection of Anthony Bryk in 2007 to replace the retiring Shulman as president of the foundation. Bryk launched an ambitious effort to "scale up" R&D in education through well-supported industrial-style prototyping and mass diffusion, beginning with an assault on the low success rates of community college students in remedial mathematics. Even as the foundation moved into new areas, including "improvement science," this signature program remained controversial among

math educators (Stigler, Givven, and Thompson 2013).

9 Outcomes assessment should be distinguished from the broader movement to increase accountability in higher education. Accountability has been linked to such performance indicators as graduation and job placement rates, and not only to learning outcomes. Performance funding, a popular approach to provide incentives for improved institutional performance, was one instrument of the broader accountability movement (see, e.g., Burke 2005; Dougherty and Natow 2009, 2015).

10 By 2001, ten states, concentrated in the South and Midwest, had experimented with or adopted standardized multiple-choice testing of student learning outcomes in publicly supported institutions (Ewell 2001b). Although the idea of demonstrating institutional value-added to learning was gaining widespread appeal, few agreed on what types of learning should be measured or how it should be demonstrated. Some advocated discipline-specific knowledge, others more general cognitive skills (such as analytical thinking and writing), and still others wanted to focus on work-related skills. Some advocated multiple-choice tests for their cost-effectiveness, but others concluded that higher-level cognitive skills could not be demonstrated in this context and required the completion of more complex, "real-world" tasks.

11 An article by the assessment expert Trudy Banta summarized the experience of educators who had attempted to implement standardized tests of general intellectual skills, such as interpretation, critical analysis, and writing. Banta argued that such instruments primarily tested entering ability; were not content neutral and therefore privileged students specializing in some disciplines more than others; contained questions and problems that did not match the learning experiences of all students at any given institution; and measured at best 30 percent of the knowledge and skills that faculty wanted students to develop. She also raised doubts about the reliability of gain scores at the individual level, the extent to which students took such tests seriously, and the dangers posed by high-stakes testing on the potential narrowing of the higher education curriculum to focus on the skills and content emphasized in the tests (Banta 2007).

12 It is important to allow for the possibility of instructor effects; it may be that better instructors are more inclined to use interactive-engagement techniques and they would also be more effective in traditional lecture formats. The consistency of the results achieved by Hake (1998) lead to the presumption that the format itself is at the very least a contributing factor to improvements in student learning.

13 The literature on quizzing is not entirely consistent. Some studies note that pre-class quizzes can contribute more to student stress than student learning if they are poorly designed or not well integrated with course learning objectives (see, e.g., Brothen and Wambach 2001). For this reason, some instructors preferred to ask students to write short answers to prompts about the reading as an alternative mechanism for ensuring that students are prepared for class.

14 Peer review of teaching has been the most common alternative to student evaluation of teaching, but peer reviewers need to be trained in evaluation and they need to have criteria and rubrics to use to evaluate their colleagues. They must then take time out of already crowded schedules to sit in on one or more classes. Without this training and these materials, peer evaluation can be as unreliable as student evaluation. Indeed, personal considerations can make it more unreliable. Colleagues in the same department may be inclined to give each

other the benefit of the doubt for the sake of departmental peace. Colleagues in adjacent departments may be more even-handed, but it is difficult to convince colleagues in adjacent departments that they have the expertise to make informed judgments about teaching outside their own discipline.

15 As Wieman and Gilbert observed: "The large observed differences in the effectiveness of difference science teaching practices and the similarity of those differences across disciplines … can be explained in terms of the basic principles of complex learning that have been established by the learning sciences. … These principles include such things as the need for intense, prolonged practice of the cognitive skills desired, with guiding feedback, and the importance of motivation and addressing the prior knowledge of the learner" (2014, 556).

16 The Teaching Practices Inventory does not begin to account for the quality of implementation of the practices it counts, a task Wieman and Gilbert consider "far more difficult" than simply assigning points for research-supported practices (2014, 561).

17 These were not the only prescriptions of the Teaching Practices Inventory. In courses that include teaching assistants, the TPI adds points for regular meetings with teaching assistants and for conversations with them about how to teach course materials. It also allocates points for what can be thought of as deep involvement in thinking about how to teach a course, including discussions with colleagues about how to teach the course, reading existing literature on approaches to teaching the course, and sitting in on the course when it is taught by colleagues.