

Ethan Hutt and Jack Schneider

A Thin Line Between Love and Hate: Educational Assessment in the United States

Educational assessment in the United States is characterized by a number of seeming contradictions. It is viewed as a valid measure of learning, but often seen as an inaccurate gauge of student ability. It is regularly the key variable in high-stakes decisions, but is widely considered to be unfair. It is used to motivate students, teachers, and schools, yet is thought to distort the learning process.

Assessment, in short, is deeply accepted and widely reviled.

How did this come to pass? Why is it that Americans are drawn to educational assessment just as strongly as they are repelled by it? What accounts for this love/hate relationship?

Our chapter seeks to unravel this seemingly contradictory stance by viewing it as the manifestation of a distinct culture — a culture with deep and particular historical roots. We believe that in order to understand educational assessment in the U.S., one must first understand the context from which views of assessment spring. What are the systems that give the American educational system its essential character? What are the core beliefs and ideologies that shape the way Americans see schooling?

Asking these questions is akin to asking about soil and the climate in one's quest to understand the plant life of a particular place. Certainly it would be possible to leap directly into a description of the plants themselves — to describe the palm trees so characteristic of Southern California, for instance, or the birch trees emblematic of the New England woods. But detailed descriptions, though they might paint a clear picture, would fail to address the basic question of why each tree grows where it does. In order to answer that, one must examine the demands of a desert environment or the qualities of a broadleaf forest.

Metaphorically, then, this chapter begins with a discussion of soil and climate — examining the cultural context of educational assessment in the United States. Only then, after establishing this basic groundwork, do we look at educational assessment itself, specifically, at the two dominant forms it has taken in the U.S.: A-F grading and standardized testing. Rather than providing a detailed history of each, we move fairly swiftly across long stretches of time, stopping to highlight particular features of grading and testing practice that evolved in response to the American context. We then conclude by outlining several common, and seemingly contradictory, beliefs

about assessment in the U.S. — beliefs that, by the end of this chapter, should be entirely understandable.

The Cultural Context of Educational Assessment

Attitudes toward educational assessment in the United States cannot be traced back to a single origin. They are the product of thousands of different factors, just as a tree is the product of its entire ecosystem.

Still, several contextual factors stand out as particularly influential. Thus, despite the fact that they represent only an incomplete subset, these systems, beliefs, and ideologies do have substantial explanatory power.

Local Control

The first essential contextual factor to consider in the American educational system is the tradition of local control, which constitutes both a structure of governance and an ideology of power.

Local control is an old tradition. States have constitutional authority over education, but from the origin of the system, power truly resided at the local level. Initially, all school funding came from local sources; and even as late as 1930, roughly 80 percent of funding came from the local level.¹ Even today, nearly half of school funding comes from cities and towns, and the laypeople elected to schoolboards exert real power over the districts they govern.

Local control is not just a structure; it is also an ideology. Consequently, there is a staunch tradition of resisting efforts — or perceived efforts — by state or federal agencies to exert control over schools in the United States.² Local discretion is fiercely guarded and fosters the belief that schools should reflect the values and concerns of their communities in everything from their approach to science curricula to their selection of common assessments.

Of course, communities are not self-governing. In the American system, state and federal authorities exert significant power, and there is general acceptance that children within each state, as well as within the nation as a whole, should be learning more or less the same thing. Even if an American educational system — a system

-
- 1 U.S. Department of Education, National Center for Education Statistics, *Digest of Educational Statistics, 2012*, accessed August 15, 2016, <http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2014015>.
 - 2 John W. Meyer and Brian Rowan, “The Structure of Educational Organizations: Environments and Organizations,” in *Environments and Organizations*, ed. Marshall W. Meyer (New York: John Wiley and Sons, 1978).

that is relatively the same for all students — does not actually exist in practice, it does exist as an idea.³

Nevertheless, centralization is strongly resisted by the whole breadth of the political spectrum, from the radical left to the conservative right. Consequently, the educational system has always had a comparatively weak infrastructure by design.⁴ When policy elites have wanted to build national systems, centralize governance, and increase efficiency, they have been largely thwarted. Instead, they have had to repurpose existing infrastructure or make common cause with private and quasi-governmental organizations to bring about change. In this context, educational assessment has often been used as a mechanism to bring an appearance of cohesion and provide more leverage for change.

Open-Access Egalitarianism

The next essential contextual factor worth considering is the idea of open-access egalitarianism.

Prior to the advent of widespread tax-supported public schools in the mid-nineteenth century, education in the U.S. was characterized by a motley assortment of private academies, dame schools, co-operative efforts, and pauper schools. No guiding ethos regarding equal access, at least beyond access to basic literacy, can be said to have existed.⁵

Funding schools through local property taxes, and later through local and state taxes, changed this.⁶ By the late 1800s, almost all students were attending tax-funded public schools. Taxpayer funding provided a structure that, in theory, ensured equal access and equal opportunity. Reality, of course, has been much more complicated. Parents angle for advantage for their children and there are wide discrepancies in funding across states, districts, and schools. Still, the vast majority of students in American schools do attend public schools.⁷

3 See, for example, works like Diane Ravitch, *The Death and Life of the Great American School System: How Testing and Choice are Undermining Education* (New York: Basic Books, 2011).

4 David F. Labaree, *Someone Has to Fail: The Zero-Sum Game of Public Schooling* (Cambridge, MA: Harvard University Press, 2012).

5 See, for example, Carl Kaestle, *Pillars of the Republic: Common Schools and American Society* (New York: Hill and Wang, 1983).

6 See, for example, Tracy Steffes, *School, Society, and State: A New Education to Govern Modern America, 1890–1940* (Chicago: University of Chicago Press, 2011).

7 U.S. Department of Education, National Center for Education Statistics, *Private School Universe Survey (PSS), 2011–12*, accessed August 15, 2016, <https://nces.ed.gov/surveys/>

As with local control, open-access egalitarianism is an ideology as much as it is an organizational feature. Critics of charter schools, for instance, often draw a false dichotomy between charters and ‘public schools’ — a rhetorical move presumably designed to impugn charters by likening them to private schools, despite the fact that they are public institutions. And in higher education, critics of public schools like the University of Michigan often disparage them as ‘private publics,’ indicating disdain for the institution’s relatively high levels of selectivity and cost, which violate the open access principle. In short, the idea of open access is not an instrumental good that produces particular outcomes; it is valued in its own right.

The belief that all students should have the same kind of educational experiences has consequences for educational assessment. On the one hand, it creates a climate conducive to testing as a means of ensuring some level of equivalence and equity across the system — a crucial concern given America’s history of segregation and discrimination. As a 2015 press release from 12 civil rights organizations put it: “Data obtained through some standardized tests are particularly important to the civil rights community because they are the only available, consistent, and objective source of data about disparities in education outcomes.”⁸ Such data, they continued, “are used to advocate for greater resource equity in schools and more fair treatment for students of color, low-income students, students with disabilities, and English learners.”⁹

On the other hand, Americans are uncomfortable with the notion that assessment systems would lead to different kinds of educational experiences for children. If assessment creates different kinds of schools — schools dominated by teaching-to-the-test, for instance, as opposed to schools where teachers exercise greater autonomy — it runs the risk of inciting backlash. Consequently, another 2015 press release — this time from the American Federation of Teachers — argued that “Testing should help inform instruction, not drive instruction [...]. We need to celebrate improvement and the joy of learning, not sanction based on high-stakes standardized tests.”¹⁰

pss/tableswhi.asp. U.S. Department of Education, National Center of Statistics, *Digest of Education Statistics 2012*, [Table 205.20](https://nces.ed.gov/programs/digest/d14/tables/dt14_205.20.asp), accessed August 15, 2016, https://nces.ed.gov/programs/digest/d14/tables/dt14_205.20.asp.

8 The Leadership Conference on Civil and Human Rights, “Civil Rights Groups: ‘We Oppose Anti-Testing Efforts,’” Press release, May 5, 2015, accessed June 5, 2015, <http://www.civilrights.org/press/2015/anti-testing-efforts.html>.

9 Ibid.

10 American Federation of Teachers, “AFT’s Weingarten on the U.S. Education Department’s ‘Testing Action Plan,’” Press release, October 24, 2014, accessed January 3, 2015,

Merit and Social Mobility

The idea that education leads to social mobility is certainly stronger today than it was 200 years ago; yet the system, from its origins, was framed as a “great equalizer.”¹¹

The idea that schools might promote social mobility is inextricably linked to the distinctly American obsession with merit. Though not an American invention, the notion of a meritocracy took fast hold in the early republic, and the nation’s founders frequently and eagerly referred to themselves as ‘Men of Merit.’ Early Americans prided themselves on their freedom from inherited ranks and titles and constructed a popular mythology about “self-made men” and an “aristocracy of talent.”¹²

If social and economic prizes are to be allocated by merit, however, there must be some means for determining the degree to which people possess it. Schools were quite convenient for this purpose. A few generations after the American Revolution, more than half of white school-age children were enrolled in school. By 1900, most young people in the northern United States, and most whites in the southern U.S. were enrolled in school. In short, for those recognized as citizens, schooling was a common experience. And schools were a natural stage for displaying merit, as a majority of young people attended, were asked to complete relatively similar tasks, and engaged in work across physical, intellectual, and moral dimensions — ostensibly a full view of a young person’s worth.¹³

Though currently debated by social scientists, the idea that schooling could lift a person up from the lowest socioeconomic ranks into the highest became a widely accepted belief by the twentieth century.¹⁴ Lawsuits across the second half of the twentieth century, for instance, and even into the twenty-first, argued for more equal educational experiences by positioning schools as a central factor in

<http://www.aft.org/press-release/afts-weingarten-us-education-departments-testing-action-plan>.

- 11 Lawrence Cremin, *The Republic and the School: Horace Mann on the Education of Free Men* (New York: Teachers College Press, 1957).
- 12 Joseph F. Kett, *Merit: The History of a Founding Ideal from the American Revolution to the Twenty-First Century* (Cornell University Press, 2012); James Bryant Conant, *Thomas Jefferson and the Development of American Public Education* (University of California Press, 1962).
- 13 Kaestle, *Pillars of the Republic*.
- 14 David F. Labaree, “Public Goods, Private Goods: The American Struggle over Educational Goals,” *American Educational Research Journal* 34, no. 1 (1997): 39–81, doi: 10.3102/00028312034001039.

determining individual social and economic outcomes. Put simply, the idea was that a level educational playing field would allow talent to shake out, and that it would create a more equal and just society — a society aligned with the founding ideals of equal opportunity and the pursuit of rational self-interest.

In order for schools to recognize merit, though, there would need to be assessments of young people, as well as a set of relatively standard measures. Whether through athletic contests, spelling bees, written examinations, or other acts of performance, young people would need to be evaluated, consistently across their peer group, in order to determine merit. And as the boundaries of such ‘peer groups’ expanded ever wider — as the nation became more connected, and as Americans became more mobile — there would be even more need for standard evaluations that would allow for comparison of students who might never actually meet face-to-face.

If assessment is to be used as a sorting mechanism — for providing differential access to rewards — it must be accepted as both fair and uniform. Yet those two demands are in conflict with each other in a massive and decentralized system characterized by high levels of student diversity. Consequently, assessment is both widely accepted and vulnerable to criticism.

Consumerism and Entrepreneurialism

The United States is a nation of hustlers — one referred to by *The Economist* as “a beacon of entrepreneurialism.”¹⁵ To use Louis Hartz’s phrase, the U.S. is an “acquisitive democracy” — a nation in which manipulating the market system for the purpose of securing material rewards is the stuff of heroism.¹⁶

This entrepreneurial spirit is not confined to the marketplace. Long before *U.S. News and World Report* began ranking high schools and colleges, parents recognized that some schools opened more doors than others and, in turn, began angling for advantage.¹⁷ Parents, for generations, have considered schooling options in the selection of towns to move to and neighborhoods to settle in. And in more extreme cases, the power of the entrepreneurial spirit drives those with

15 “The United States of Entrepreneurs,” *The Economist*, March 12, 2009, accessed May 24, 2015, <http://www.economist.com/node/13216037>; see also Labaree, *Someone Has to Fail*.

16 Louis Hartz, *The Liberal Tradition in America* (New York: Harcourt Brace and Co. 1955), 138.

17 David F. Labaree, *The Making of an American High School: The Credentials Market and the Central High School of Philadelphia, 1838–1939* (New Haven: Yale University Press. 1988).

means to venture *outside* the public system — to private schools that maintain strong connections with prestigious colleges and elite firms.¹⁸

On its face, the ethos of consumerist entrepreneurialism seems directly to contradict the ideology of open-access egalitarianism. Yet it is often the case that those working the system to provide their children with a presumed advantage still believe strongly in open access. They are not acting in accordance with what they wish the system would become. Rather, in the words of David Labaree, “consumers are simply pursuing their own interests through the medium of education [...] they’re just trying to get ahead or at least not fall behind.”¹⁹

This duality can be further explained by dividing the utility of education into public goods and private goods. As a public good, the content of education matters a great deal. All members of society benefit from an education system that serves all students equally well — preparing citizens for the republic and employees for the capitalist economy. As a private good, however, the content of education matters much less. “Front and center in the consumer agenda for gaining the greatest benefit from schooling,” writes Labaree, “is to acquire its marketable tokens of accomplishment. These include gold stars, test scores, grades, track placements, academic credits, and — most of all — diplomas.”²⁰

Educational assessment, then, matters a great deal to entrepreneurial consumers of education. If grades and test scores are a currency, then savvy parents will seek to acquire the currency of greatest value and to get as much of it as possible. They will make calculated investments, negotiate whenever possible, and leverage any advantage.

Populism and Common Sense

The United States is a nation infatuated with common sense solutions to knotty problems. From the country’s origins, a kind of anti-intellectualism has been deeply ingrained in national culture. Popular beliefs, as evidenced through long-standing deference to leaders of business and evangelical religion, often hold greater sway — at least in the public realm — than the theories of experts. In fact, Americans have a distinctly derisive word for experts. As Richard Hofstadter wrote in 1963, “the greater part of the public...has an ingrained distrust

18 See, for instance, Shamus Rahman Khan, *Privilege: The Making of an Adolescent Elite at St. Paul’s School* (Princeton, NJ: Princeton University Press, 2012).

19 Labaree, *Someone Has to Fail*, 236.

20 *Ibid.*, 237.

of eggheads.”²¹ Divorced from reality, isolated in ivory towers, and wedded to unwieldy theories, eggheads cannot see the forest for the trees. They overlook the simple solution obvious to the attentive layman.

This is not to say that Americans reject expertise. Far from it. There is a great deal of deference to experts. But expertise is always open to challenge, and common sense is an accepted evidence base for even the most complicated policy proposals.

In schooling, particularly, this has long been the case.²² Consider teachers, who are the experts within classrooms, but who have notoriously low levels of prestige given the common assumption that anyone with common sense can teach.²³ Consider, too, how school districts are governed — by boards of elected laypeople, often with little expertise beyond their own school experiences. Or consider the sway that local communities have maintained over the curriculum — rejecting the teaching of particular subjects, like evolution or sex education, or rejecting particular curricular narratives, like those related to historically-rooted inequality.

Americans, thus, feel qualified to weigh in on nearly any issue in social and political life. But perhaps particularly so in education, which they have all experienced for often well over a decade. When allied with politicians who favor populist approaches to policy — as evident in the recent push to roll back the Common Core State Standards — or with deep-pocketed foundations who favor common sense over evidence, the ideology of anti-intellectualism can exert a powerful influence.

With particular regard to educational assessment, populist common sense has been notably evident. Over the decades, assessment has transformed into highly complex work, often impenetrable to laypeople. Consequently, the pushback against ever more opaque assessment practices has often taken the explicit form not only of popular outrage at particular assessment devices, but, increasingly, at the very practice of assessment itself. In short, Americans are willing to defer to testing experts; but the increasingly inscrutable practices of these experts also makes them suspect in the eyes of populists.

21 Richard Hofstadter, *Anti-Intellectualism in American Life* (New York: Knopf, 1963).

22 The value of common sense on school matters remains an important concern for contemporary reformers. See, Jack Schneider, *Excellence For All: How a New Breed of Reformers Is Transforming America's Public Schools* (Nashville, TN: Vanderbilt University Press, 2011); David Labaree, *The Trouble with Ed Schools* (New Haven: Yale University Press, 2004).

23 Dan Lortie, *Schoolteacher: A Sociological Study* (Chicago: University of Chicago Press, 1975).

American Culture and Practice of Assessment

Having laid out the climatic conditions that give rise to a particular culture of assessment, we think it instructive to examine the development of the two most common assessment practices in the United States. Though space does not permit a full historical accounting, we have selected key developments in the evolution of these practices that illustrate the interaction between cultural context and assessment practice.²⁴

A Brief History of Grading

There are few more central practices in American schooling than grading. Assignments are graded, participation in graded, projects are graded, tests are graded. So central are grades to students' schooling experience, in fact, that they frequently become a part of students' personal identity. 'She's a straight A student' is a phrase intended to convey not only a general description of a student's academic record, but also a more general statement about the *kind* of person the student is: conscientious and hard-working — a striver. And yet, as much as grades have become a ubiquitous part of American schooling, society, and student identity, they have long been reviled as, at best, incomplete and, at worst, corrosive to educational aims. Their evolution and steady spread despite these critiques, then, offers an instructive example for thinking about the dynamics of assessment in the U.S.

At the core of the history of American grading are two sets of tensions. The first is between the internal and external organizational demands of grading — the need for grading to communicate clearly within a school versus the need for grading to communicate something beyond and across schools. The second tension, interwoven with the first, is between the particularistic, local meaning of grades and the standardized, universal meaning of grades.

These tensions can usefully be thought of as reflecting various combinations of the contextual factors identified at the outset of this chapter. For example, the demand for particularized meanings of grades, comporting with both family and teacher notions of local control and discretion, run directly counter to the desires of students seeking wider recognition of their academic merit, as well as to the desires of policymakers seeking egalitarianism through standardization. Efforts

24 For a full history of the development of American grading practices see Jack Schneider and Ethan Hutt "Making the grade: a history of the A–F marking scheme," *Journal of Curriculum Studies* 46, 2 (2014): 201–24, doi: 10.1080/00220272.2013.790480.

to balance such tensions, as this section will demonstrate, have led to constantly shifting practices within a more broadly stable culture of assessment.

The early days of grading were characterized by local control. Idiosyncratic practices reflected the nominally professional vision of individual teachers and the communities that banded together to hire them. These practices were inherently ephemeral, as early evaluation in American common schools usually took the form of oral examinations and public performances.²⁵ Far from being recorded as part of the permanent record, the assessment was intended to be both instantaneous and dynamic: students received instant praise (or rebuke) and often saw their physical location in the class reflect this regular assessment. Students who did well were moved toward the front of the class, while other students found themselves moving figuratively and literally behind their peers. These kinds of assessment practices were well suited to the early 19th century model of American schooling — one characterized by irregular student attendance, varied curricula, and the one-room school house.

As the 19th century progressed, American educators, influenced by the practices of their colleagues abroad, increasingly began to worry about the messages these grading practices sent to students — placing an emphasis as they did on the immediate goal of ‘winning’ the competition rather than on the long-term goal of educational self-improvement. As famed educational reformer Horace Mann observed, “if superior rank at recitation be the object, then, as soon as that superiority is obtained, the spring of desire and effort for that occasion relaxes.”²⁶ Policy leaders like Mann believed that replacing daily, ephemeral, performative competition with periodic written, private records of student achievement — like the report card — would turn student attention back to the intrinsic, long-term goals of education and the persistent effort their achievement required. Yet though the report card invited students to take a longer-view of their education, it did not prevent them from bending these practices to their own entrepreneurial ends.

After the introduction of innovations like the report card for recording grades and formally communicating them to students and parents, the form and substance of written grading practices developed substantially over the last decades of the 19th century, reflecting the shifting organizational forms of American schooling. Beginning first with American colleges and disseminating downward through the system, American educators increasingly marked school progress not in years

25 William J. Reese, *Testing Wars in the Public Schools: A Forgotten History* (Cambridge, MA: Harvard University Press, 2013).

26 Horace Mann, *Ninth Annual Report* (Boston, MA: Dutton and Wentworth, State Printer, 1846), 504.

but in grade levels and course credit — reflecting the increasing number of curricular options and tracks available to students. One upshot of this was that it was no longer possible for schools to rank students directly with each other because they were not taking the same courses. This created pressure to recognize students as belonging to *categories* defined by their grades and, in higher education, by their graduation honors (e.g. cum laude, summa cum laude). Here again, reformers hoped that these categories would displace competition for individual honors in favor of groups of men with equal distinction. But the value to individuals in achieving these ranks could only reduce competition so much — it left plenty of room for ambitious strivers to make calculated choices for the purpose of achieving the marks that would allow them to get ahead.

Concurrently, schools at all levels began to narrow the meaning of grades in order to communicate more clearly their academic distinction. For much of the 19th century, grades reflected both a moral and behavioral dimension, as well as an academic one; but in the second half of the century, grades came to reflect only a student's academic performance.²⁷ This move reflected the increasing academic focus of schools, and the increasing importance of academic grades as legitimate markers of distinction beyond the confines of the school. As the president of Philadelphia's celebrated Central High School explained, "the best scholars are not always the most decorous;" consequently, combining behavior and achievement only served to "depriv[e] the student of those honors which he had fairly won by diligence and industry."²⁸

It is only a slight simplification to say that the increasing association of grades with academic merit, and of academic merit with social distinction, created increasing pressure to standardize grades. After all, if grades were to work for the purpose of open access egalitarianism, they needed to have uniform meaning across levels of the school system and throughout the nation. And though this ran against the tradition of local control, only so much local variation could be tolerated in an increasingly national competition for academic merit and social mobility.

Greater uniformity in grading would place serious constraints on the use of grades as contextualized communication tools — as a way of indicating a student's standing within the class. And there was strong pressure for greater uniformity. As one early 20th century commentator put it:

27 Labaree, *The Making of an American High School*; Frederick Rudolph, *The American College and University: A History* (Athens, GA: University of Georgia Press, 1962).

28 Labaree, *The Making of an American High School*, 141.

“the one common language in which the scholarly attainments of pupils are expressed is a scalar one [...] if we, who live in the Middle West, read in a New York Magazine that a certain man entered college with an average grade of 95 [...] we know pretty well what that means; and so it is in the country over.”²⁹

Yet despite the universalistic meaning of these grades, there was no obvious method “whereby grades assigned by one teacher can be intelligently compared with those assigned by another, and all brought to a common standard.”³⁰ Another contemporary commentator put the matter — and stakes attached — even more plainly: “the grade has in more than one sense a cash value and if there is no uniformity of grading in an institution, this means directly that values are stolen from some and undeservedly presented to others.”³¹

These calls to shore up faith in grades as the uniform currency of academic exchange resonated well with the ascendant system-building push of Progressive Era educators. This new generation of school reformers was actively interested in using the tools of government and emerging educational science to bring more coherence, higher standards, and greater efficiency to America’s sprawling education system.³² With this systemization came greater attention and a policy push to standardize teacher grading practices — to ensure that an “A” in Portland, Maine, for instance, would have the same meaning as an “A” in Portland, Oregon. A number of studies published in this period made a great deal about the unreliability and subjectivity of teacher grading and the frequent failure of teacher grades to adhere to characteristics of the normal curve, which had become such a touchstone for psychologists and others working within the mental testing movement.³³

While the majority of school officials by the mid-20th century accepted the A-F format for grading, as well as the premise of a standardized grade distribution, progress was uneven and contested. Many educators wondered openly about the role of the teacher’s own professional discretion in the face of these constraints, pleading for the legitimation of an alternative to the normal curve — a format

29 Leroy D. Weld, “A Standard of Interpretation of Numerical Grades,” *The School Review* 25, no. 6 (1917), 412.

30 *Ibid.*

31 Max Meyer, “Experiences with the Grading System of the University of Missouri,” *Science* 33, no. 852 (1911), 661.

32 David Tyack, *The One Best System: A History of American Urban Education* (Cambridge, MA: Harvard University Press, 1974); Raymond Callahan, *Education and the Cult of Efficiency* (Chicago: University of Chicago Press, 1964).

33 See, for instance, Isador E. Finkelstein, *The Marking System in Theory and Practice* (Baltimore, MD: Warwick and York, 1913).

that might give individual teachers more discretion.³⁴ Others rejected the premise of reducing student work to a single grade by opting to substitute the report card for a letter home or attempting to efface their meaning and power by reducing the number of categories to Pass/Fail or High Pass/Pass/Fail. Still, these efforts never amounted to much more than rear guard actions against the continued spread and use of grades.

Indeed, the reasons that these practices were rejected provide important insight into the American culture of grading. While many educators and parents complained, as they had in the 19th century, that competition around academic distinction encouraged students to seek grades rather than knowledge, the increasing embeddedness of grades beyond the school environment made it difficult to get out from under their use.

Even when taken on purely academic terms, those who critiqued grades often chose to operate within their basic framework, even as they sought greater distinction, and social mobility, through grades. Concerned that students could have identical grade point averages despite very different academic course selections, the parents of high achieving students, and the colleges that sought their enrollments, pushed for ‘weighted-GPAs,’ with students receiving extra credit for harder courses. Likewise, even students at schools with alternative grading practices worried that these practices might limit their prospects in subsequent academic and labor market settings. For example, the University of California Santa Cruz, founded as a radical alternative to existing state institutions, recently abandoned its three decade-old practice of giving students “narrative evaluations” in lieu of letter grades, citing concerns about student graduate school competitiveness and the inefficiency of narrative grades when done at large-scale.³⁵ The value of long-form communication between teacher and student, it seems, had to give way to the practical considerations of mass higher education and international labor markets.

As this brief history of grading should make clear, the set of considerations shaping assessment practices extend far beyond those of the teacher in the classroom or the family of the student. Rather, these practices, like schools themselves, are embedded in a much larger context that establishes both demands and constraints.

34 Norman E. Rutt, “Grades and Distributions,” *National Mathematics Magazine* 18, no. 3 (1943): 120–26.

35 Tanya Schevitz, “UC Santa Cruz Faculty to Vote on Ending Narrative Evaluations,” *San Francisco Chronicle*, November 26, 2000, accessed July 15, 2014, <http://www.sfgate.com/education/article/UC-Santa-Cruz-Faculty-to-Vote-on-Ending-Narrative-2726032.php>.

Grades, then, were transformed into a source of academic distinction for students — trophies of their success in open academic competition — and a currency with a reasonably stable, if not entirely certain, external value. The failure to fulfill both of these roles served as a persistent spur to move in the direction of greater standardization, competition, and distinction. Moreover, such tensions led to a further elaboration of grading practice, rather than to a rejection of the underlying culture of assessment.

A Brief History of Standardized Testing

As with the history of grading, the history of standardized testing has been shaped by the interplay between cultural factors and the specific ends to which standardized testing has been put. Yet though both are clearly part of the assessment culture of American schooling, there are inherent differences between grading and testing practices — differences instructive for showing how the same dynamic can drive the development of a very different set of practices.

Specifically, while the history of grading represents an attempt to make a personal communication between teacher and student more durable and intelligible to people at increasingly greater remove — both spatially and temporally — standardized testing was always fundamentally about creating standardized comparisons and the creation of larger ‘communities.’ As with grading, standardization of testing practice provided both an opportunity and a threat. In drawing larger numbers of students together from across real communities, standardized tests offered students a chance to garner greater distinction in an even larger academic competition. In doing so, however, they threatened existing hierarchies, conventional wisdom, and other ways of knowing about schools. Thus, if the challenge of developing grading practices was to maintain the personal dimension of grades as they became increasingly standardized, the challenge of standardized testing was to maintain the simplicity of common metrics amid a blizzard of competing and conflicting test scores.

The first written standardized test was introduced into American public schools in Boston in 1845. Even in this very first instance, the dynamics that would animate the future use and proliferation of standardized tests were easy to see. As with the introduction of recorded grades, standardized tests were ushered into schools in an effort to replace impressionistic evaluations of students with ‘hard facts.’ Unlike recorded grades, however, the introduction of tests was done with the explicit intention of engaging the larger public in a debate about school quality. The men responsible for introducing standardized tests to Boston, including

none other than Horace Mann, believed that the (low) test scores would provide evidence of the need to adopt their favored school reforms.³⁶

While reformers succeeded in calling into question the presumed superiority of Boston schools and pushing through some of their desired reforms, they did not count on the extent to which these new tools would be taken up by others. It did not take long, for instance, before publishing companies produced an avalanche of test materials for educators and test preparation books for students. And though rote teaching was an explicit target of the reformers who introduced standardized testing, future generations agreed that tests had done as much to solidify such teaching practice as they had to displace it. These consequences reflect the tensions between merit and entrepreneurialism, on the one hand, and between merit and equal access egalitarianism on the other. Reformers seeking to inform the public about undesirable variation and leverage their civic pride in having widely renowned schools were increasingly surprised, then, as they encountered an entrepreneurial perspective that cast testing not an opportunity for reform, but as a chance to get ahead.

Over the next half century, tests would come to occupy an increasingly prominent place in the organization and operation of schools. Individual schools utilized tests and norm tables to determine student placement; districts used standardized tests to allocate the limited number of seats in high schools; and states like New York introduced a state wide examination — the Regents Examination — intended to standardize achievement norms and direct competitive energies toward higher academic standards.³⁷

The entrepreneurial and meritocratic dimensions of schooling were well-served by the introduction of standardized tests. But so was open-access egalitarianism. As in the case of grades and grading, those who had a vested interest in academic achievement had an interest in making academic merit universally recognizable. By allowing more direct and far-flung comparisons, standardized tests were, from the beginning, enlisted in this cause. As a New York educator explaining the value of a diploma backed by the faith and credit of a statewide standardized test observed to his students, “we may think, and you may believe that what you do here is recognized throughout the state, but outside of our village we are not

36 Reese, *Testing Wars in the Public Schools*.

37 Stephen Jay Gould, *The Mismeasure of Man* (WW Norton & Company, 1996); Labaree, *The Making of an American High School*; Nancy Beadie, “From Student Markets to Credential Markets: The Creation of the Regents Examination System in New York State, 1864–1890,” *History of Education Quarterly* 39, no. 1 (1999): 1–30. doi: 10.2307/369330.

known; your papers simply show that you have got a certificate somewhere.”³⁸ Such sentiments became increasingly common among the self-styled professional educators and purveyors of the educational sciences during the Progressive Era. As one commentator explained: “Test scores furnish the common language, for anyone can understand what is meant by saying that our schools in Smithville are a year ahead of most schools of America in, say, arithmetic, and a year or two years behind others in music or French or manual training.”³⁹

Such comparisons could form the basis for arguments to improve schools. As one set of Progressive Educators advised superintendents: “Test results constitute incontrovertible facts, so often needed by the superintendent in a campaign of education of public opinion.”⁴⁰ But they could also provide an unwelcome entanglement for those who fared poorly. As another commentator of the period observed: “If one could read all the small town papers of any given state for one year, he would probably find three-fourths of them claiming that their home town had the best schools in the state.”⁴¹ Put simply, the tradition of local control, particularly when combined with civic pride, militated against the uncomplicated acceptance of test results.

American educators may have embraced standardized testing but, true to form, they did not do so in any uniform way. The sprawling American education system begat a standardized testing industry that was every bit as eclectic and varied as the system it served and helped to create. By one estimate, there were at least 250 commercially available standardized tests in 1922. The sheer variety of tests ensured that there would be no definitive answer to the best school district, best instructor, or the best method of math instruction because competing claims could easily be supported by their own respective set of standardized test scores.

Unlike Europe, where the responsibility to commission and oversee high stakes qualifying examinations fell to the state, these responsibilities were outsourced to independent entities in the U.S. — sometimes to non-profit organizations, but just as often to for-profit companies. Thus, even in college admissions — an area of comparatively strong standardization — two separate organizations and tests were allowed to develop and serve an indispensable role in college access.

38 Beadie, “From Student Markets to Credential Markets,” 24.

39 Denton L. Geyer, *Introduction to the Use of Standardized Tests* (Chicago: Plymouth Press), 11–12.

40 Sidney L. Pressey and Luella Cole, *Introduction to the Use of Standard Tests: A Brief Manual in the Use of Tests of Both Ability and Achievement in the School Subjects* (New York: World Book Company, 1922), 34–35.

41 Carter Alexander, *School Statistics and Publicity* (Boston: Silver, Burdett), 52–53.

To this day, both the Scholastic Assessment Test (SAT)⁴² and American College Test (ACT), emphasize their different approaches to measuring college readiness, and enterprising students are routinely given the advice to take both and submit their highest score.

There was also rarely any incentive for any of the parties involved to commensurate the values or scores produced by the multitude of tests. It was better to leave these self-contained standardized visions to coexist: testing companies had market share to protect; students had records to burnish; and districts (and cities) had political, economic, and civic interests to project an image of scholastic excellence and efficiency. In many cases this permissiveness was a double-edged sword. On the one hand, it allowed those without privilege to move up through the system. On the other hand, however, it allowed their more privileged peers to modify or abandon those metrics which no longer served their interests.

This struggle to enlist standardized test in non-uniform aims has reached its fullest expression in the era of federally-mandated high-stakes testing, which began when No Child Left Behind was signed into law by George W. Bush in 2002. Despite the explicit attempt to use standardized test scores to promote equity and allow for easier comparison of schools, districts, and states, the result produced not one clear fact sheet, but rather, a 50-hued tie-dye of competing tests, standards, and definitions of proficiency. Though many lament this development as the unfortunate byproduct of federalism, we think it is better understood as the logical result of the competing ends that Americans seek to reach through standardized testing. Such a system accommodates the contradictory aims of egalitarianism and advantage, localism and universality.

Just as importantly, for those who reject the premise of quantifying learning — a line of critique that runs uninterrupted through American educational history — the lack of clarity produced by the blizzard of test scores only provides further evidence that the experts and their tests misapprehend what is *really* going on in schools. This is the basis of the current movement by parents to ‘opt-out’ of state and federal mandated testing. The message from parents could not be clearer: test results dissemble; they have nothing to offer a parent in the know. And yet, despite the outcry against the ‘over-testing’ of students, the U.S. Congress balked at the chance to remove testing requirements from the latest version of federal law, now rechristened the ‘Every Student Succeeds Act’. Instead, they offered states more flexibility, more discretion in the design, use, and

42 Originally called Scholastic Aptitude Test.

oversight of tests — ensuring that tests would continue to balance the multiple demands to which they have always been put.

Contradictory Beliefs about Assessment

In the United States, educational assessment is characterized by deep acceptance and persistent complaining. It is remarkably stable, but constantly in a process of churn.

This is the result of contradictory beliefs about assessment — beliefs rooted in a particular cultural context — that simultaneously pull American schools in different directions. And given this rooting, assessment can never be exactly what Americans want it to be. Their feelings must always be mixed.

In this section, we will take a closer look at several examples of seemingly contradictory notions about assessment that are common in the U.S. These are not intended to be comprehensive; they do not sum up all that Americans believe about testing. However, these examples are representative of a distinct ethos. And, collectively, they constitute a long stretch of the thin line this paper seeks to map out — the line between love and hate that describes the American culture of educational assessment.

Notion 1: Assessment systems should be uniform and coherent, but not national.

Open-access egalitarianism requires a single system that treats all people fairly. And the impulses of merit and social mobility demand that all students take the same tests, opening the same opportunities to receive rewards in exchange for their talent and hard work. The result has been the creation and utilization of a variety of national assessment systems like the SAT, the ACT, the AP (Advanced Placement Program), the NAEP (National Assessment of Educational Progress), and the PARCC (Partnership for Assessment of Readiness for College and Careers).

Yet there is pushback against these assessment systems because each, in some way, erodes local control. Additionally, any national assessment threatens consumer autonomy in a free market by limiting available choices. Consequently, national assessment systems in the United States are always something short of national. In the case of assessments, the SAT, ACT, and AP tests (Advanced Placement tests), for instance, students can opt-in and opt-out of the exams at will. And in the case of the federally-mandated NAEP exam, the test may be national, but it is also toothless. With no stakes attached to the test, students have nothing

to lose by taking it. And without results aggregated at the state level, rather than the student or school level, it is not exactly clear what, exactly, NAEP assessing.

When no compromise is evident, as was the case with the PARCC exam — a test tied to the Common Core State Standards project that sought to unify the nation's 50 different curricular and assessment systems — hostility can build up quickly. Despite the fact that the PARCC exam represented only a tweak to the system, the perception that it was forcing standardization at the national level produced a surprising coalition of opponents. Overwhelming support in favor of PARCC quickly reversed.

Notion 2: Assessments should create winners, but not losers.

Because of the pressure for open-access egalitarianism, there is consistent demand for assessments that produce comparable information. All students, for instance, should be given grades that have relatively equal meaning and should all sit for the same exams. Additionally, information gathered from such assessments should provide the basis for assuring that students have equal opportunities to advance in a meritocratic system.

Yet the forces of consumerism and entrepreneurship mean that individual families, seeing education as a private good, are generally unwilling to accept that their children have lost in a meritocratic competition. In a system with a single high-stakes national test — as in Korea or Japan — this could be problematic. But in the permissive American assessment culture, these impulses are channeled into the cultivation of new and different forms of distinction. AP tests, for instance, award college credit, but are optional, and students do not suffer for earning low marks on them — they can choose, as with SAT scores, whether and which scores to report colleges. And students often shop for the easiest AP tests to take. In this light, standardized assessment becomes a means not of selection and exclusion, but of building an argument and making a case for one's individual merit.

Grade inflation is perhaps the clearest example of the contradictory desire to produce winners but not losers. While students are graded on a uniform A-F scale, parents and students exert strong pressure on teachers to give good grades. And even if pressure is not applied directly, teachers know that grades can seriously impact a student's advancement through the system. Not surprisingly, then, the modal grade at Harvard University is an A- and the second most frequently issued grade is an A. This compression of grades at the top end of the scale is indicative of the desire to keep as many students in the game as possible.

Notion 3: Assessment results are objective, but not always valid.

The default position in the United States is that the use of educational assessments is a fair and objective way to measure student ability. This is particularly true for winners in the system, who have a vested interest in showing that their advancement is due to merit rather than chance or privilege. The numbers don't lie.

But when the drive for merit-based social mobility is thwarted, it tends to engender populist pushback against assessment developers who, detractors argue, do not understand what really goes on in schools or what it really means to be educated. One clear example of this can be seen in the history of the AP program. Elite high schools championed the AP program, which gave their students an admissions edge. But when the program expanded more broadly, elite schools, which no longer enjoyed an admissions advantage from AP, began dropping the program. They criticized AP as out-of-date and out of touch — an invalid measure of advanced work. Subsequently, many of these schools entrepreneurially crafted alternatives that they branded “post AP” curricula.⁴³

Another clear example can be seen in the pushback against state assessment systems adopted under No Child Left Behind. As long as test results did not adversely affect students and their schools, the law garnered widespread support. But when schools began closing, and when researchers began documenting unintended consequences like narrowing of the curriculum and teaching-to-the-test, critics began arguing that the tests were invalid and not worth the costs associated with them.

Notion 4: Assessments should have stakes, but should not distort the process of learning.

Because K-12 education is taxpayer funded, Americans tend to support the idea that public schools should produce results. As they see it, students should not simply drift through the system, learning or not learning according to chance or desire. From this perspective, assessments should be tied to accountability mechanisms as a way of promoting effective management at the local and state level.

Additionally, parents want information about their children's progress, and support mechanisms that help them encourage effort. Given belief in the relatively meritocratic nature of an open-access system, parents believe that grades and test

43 Jack Schneider, “Privilege, Equity, and the Advanced Placement Program: Tug of War,” *Journal of Curriculum Studies* 41, no. 6 (2009): 813–31; Jack Schneider, *Excellence For All*.

scores provide tools that will help them guide their children toward social mobility. Assessment results are useful carrots and sticks.

Yet Americans do not want assessments to serve as ends in and of themselves. Assessment, they generally believe, should not distort the learning process. And it should especially not do so in a differential way — affecting some schools and students, but not others.

State accountability tests, adopted in the wake of No Child Left Behind, were designed to motivate teachers and students to focus more on the acquisition of content. But the strongest criticism has been that such tests have distorted the learning process — turning too much curricular attention to tested subjects, and orienting schools too strongly toward test results. Consequently, a growing ‘opt-out’ movement has sprung up, with parents exempting their children from the standardized tests that continue to be mandated by federal policy — a federal policy with widespread support.

Conclusion

In this chapter we have tried to sketch out an answer to a fundamental question: How is it that Americans both embrace and revile educational assessment?

As we suggest, the answer lies deep below what is immediately visible in present practices. Thus, any attempt to unravel this mystery must begin with an examination of the conditions that have given shape to the way Americans approach assessment.

At the outset of this chapter, we used the metaphor of soil and climate, making the case for the importance of context. In the specific case of educational assessment in the United States, however, a better metaphor may be that of tectonic plates. Invisible above the surface, the forces of consumerism and entrepreneurialism, merit and social mobility, open-access egalitarianism, and local control grind constantly against each other. Exerting pressure in different directions, they continue to shape the landscape above.

At various times, different combinations of these values have inspired the pursuit of seemingly contradictory aims. Yet contradictory as they may seem, these aims are consistent with the cultural forces that continue to shape both belief and action. Activity above ground, in other words, aligns with action below.

The resulting picture of assessment in the United States, then, is one of a fault line under intense pressure. This pressure is frequently relieved in small adjustments — the removal of a contested metric, for instance, or the development of a new test. And occasionally, it manifests in disruptive ground-shaking. But the

fault line itself never dissolves. Below ground, the forces of love and hate grind on. They do so in equal measure.

Literature

- Alexander, Carter. School Statistics and Publicity.* Boston: Silver, Burdett, 1919.
- American Federation of Teachers.* "AFT's Weingarten on the U.S. Education Department's 'Testing Action Plan.'" Press release, October 24, 2014. Accessed January 3, 2015. <http://www.aft.org/press-release/afts-weingarten-us-education-departments-testing-action-plan>.
- Beadie, Nancy.* "From Student Markets to Credential Markets: The Creation of the Regents Examination System in New York State, 1864–1890." *History of Education Quarterly* 39, no. 1 (1999): 1–30. doi: 10.2307/369330.
- Callahan, Raymond.* *Education and the Cult of Efficiency.* Chicago: University of Chicago Press, 1964.
- Conant, James Bryant.* *Thomas Jefferson and the Development of American Public Education.* University of California Press, 1962.
- Cremin, Lawrence.* *The Republic and the School: Horace Mann on the Education of Free Men.* New York: Teachers College Press, 1957.
- Finkelstein, Isador E.* *The Marking System in Theory and Practice.* Baltimore, MD: Warwick and York, 1913.
- Geyer, Denton L.* *Introduction to the Use of Standardized Tests.* Chicago: Plymouth Press, 1922.
- Gould, Stephen Jay.* *The Mismeasure of Man.* New York, London: WW Norton & Company, 1996.
- Hartz, Louis.* *The Liberal Tradition in America.* New York: Harcourt Brace and Co. 1955.
- Hofstadter, Richard.* *Anti-Intellectualism in American Life.* New York: Knopf, 1963.
- Kaestle, Carl.* *Pillars of the Republic: Common Schools and American Society.* New York: Hill and Wang, 1983.
- Karabel, Jerome.* *The Chosen: The Hidden History of Admission and Exclusion at Harvard, Yale, and Princeton.* New York: Houghton Mifflin Harcourt, 2006.
- Khan, Shamus Rahman.* *Privilege: The Making of an Adolescent Elite at St. Paul's School.* Princeton, NJ: Princeton University Press, 2012.
- Kett, Joseph F.* *Merit. The History of a Founding Ideal from the American Revolution to the Twenty-First Century.* Cornell University Press, 2012.

- Labaree, David F. *The Making of an American High School: The Credentials Market and the Central High School of Philadelphia, 1838–1939*. New Haven: Yale University Press, 1988.
- . “Public Goods, Private Goods: The American Struggle over Educational Goals.” *American Educational Research Journal* 34, no. 1 (1997): 39–81. doi: 10.3102/00028312034001039.
- . *The Trouble with Ed Schools*. New Haven: Yale University Press, 2004.
- . *Someone Has to Fail: The Zero-Sum Game of Public Schooling*. Cambridge, MA: Harvard University Press, 2012.
- Lortie, Dan. *Schoolteacher: A Sociological Study*. Chicago: University of Chicago Press, 1975.
- Mann, Horace. *Ninth Annual Report*. Boston, MA: Dutton and Wentworth, State Printer, 1846.
- Meyer, Max. “Experiences with the Grading System of the University of Missouri.” *Science* 33, no. 852 (1911): 661–67.
- Meyer, John W., and Brian Rowan. “The Structure of Educational Organizations: Environments and Organizations.” In *Environments and Organizations*, edited by Marshall W. Meyer, 78–109. New York: John Wiley and Sons, 1978.
- Pressey, Sidney L., and Luella Cole. *Introduction to the Use of Standard Tests: A Brief Manual in the Use of Tests of Both Ability and Achievement in the School Subjects*. New York: World Book Company, 1922.
- Ravitch, Diane. *The Death and Life of the Great American School System: How Testing and Choice are Undermining Education*. New York: Basic Books, 2011.
- Reese, William J. *Testing Wars in the Public Schools: A Forgotten History*. Cambridge, MA: Harvard University Press, 2013.
- Rudolph, Frederick. *The American College and University: A History*. Athens, GA: University of Georgia Press, 1962.
- Rutt, Norman E. “Grades and Distributions.” *National Mathematics Magazine* 18, no. 3 (1943): 120–26.
- Schevitz, Tanya. “UC Santa Cruz Faculty to Vote on Ending Narrative Evaluations.” *San Francisco Chronicle*. November 26, 2000. Accessed July 15, 2014. <http://www.sfgate.com/education/article/UC-Santa-Cruz-Faculty-to-Vote-on-Ending-Narrative-2726032.php>.
- Schneider, Jack. “Privilege, Equity, and the Advanced Placement Program: Tug of War.” *Journal of Curriculum Studies* 41, no. 6 (2009): 813–31. doi: 10.1080/00220270802713613.
- . *Excellence For All: How a New Breed of Reformers Is Transforming America’s Public Schools*. Nashville, TN: Vanderbilt University Press, 2011.

- Schneider, Jack, and Ethan Hutt. "Making the grade: a history of the A-F marking scheme." *Journal of Curriculum Studies* 46, 2 (2014): 201–24. doi: 10.1080/00220272.2013.790480.
- Steffes, Tracy. *School, Society, and State: A New Education to Govern Modern America, 1890–1940*. Chicago: University of Chicago Press, 2011.
- The Leadership Conference on Civil and Human Rights*. "Civil Rights Groups: 'We Oppose Anti-Testing Efforts.'" Press release, May 5, 2015. Accessed June 5, 2015. <http://www.civilrights.org/press/2015/anti-testing-efforts.html>.
- "The United States of Entrepreneurs." *The Economist*, March 12, 2009. Accessed May 24, 2015. <http://www.economist.com/node/13216037>.
- Tyack, David. *The One Best System: A History of American Urban Education*. Cambridge, MA: Harvard University Press, 1974.
- U.S. Department of Education, National Center for Education Statistics, *Digest of Educational Statistics, 2012*. Accessed August 15, 2016. <http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2014015>.
- . *Digest of Education Statistics 2012, Table 205.20*. Accessed August 15, 2016. https://nces.ed.gov/programs/digest/d14/tables/dt14_205.20.asp.
- . *Private School Universe Survey (PSS), 2011–12*. Accessed August 15, 2016. <https://nces.ed.gov/surveys/pss/tableswhi.asp>.
- Weld, Leroy D. "A Standard of Interpretation of Numerical Grades." *The School Review* 25, no. 6 (1917): 412–21.