CHAPTER 2

A Brief History of Grading

In this chapter, we answer the following questions:

1. *What were significant societal trends and beliefs in the United States during the first half of the twentieth century, and how were they manifested in schools?*
2. *What were the original purposes and designs of our current grading practices?*

♦ ♦ ♦

> *Grading remains a central feature of nearly every student's [and teacher's, and parent's] school experience. As such, it can be easy to perceive them as both fixed and inevitable—without origin or evolution. An effect of this is that despite their limitations, grades are often accepted quite uncritically by all parties involved.*
>
> *(Schneider & Hutt, 2014)*

Grading is part of the "grammar of schools" (Tyack & Cuban, 1995), a concept so embedded in our idea of what a school is that it seems silly to question it. What seems more foundational to everyone's school experience than getting grades?

Of course, the ways we grade weren't handed down from heaven, but they have an origin and an evolution. It turns out that our current grading practices were designed over 100 years ago, reflecting and responding to the needs of the United States in the early twentieth century, and intending to solve problems of that context. What is surprising is that many of those original designs haven't changed since. Our approaches to teaching and learning have radically shifted, but our grading practices have been preserved in amber, and we continue to use them reflexively and often uncritically. As a first step toward more informed decisions about grading, we must examine and excavate the history of grading. Of all possible approaches to grading, why do we have the ones we have? What were the larger ideas in the American political, economic, and social context of the early twentieth century, and how were

schools, and their grading practices, a manifestation of those ideas? What were needs and problems that schools generally, and grades specifically, were intended to solve? Who defined those needs and problems, and what were deemed appropriate solutions? While what follows is not intended to be a definitive history of grading and the broader history of schools, if we're going to understand our grading, question it, and find ways to improve it, particularly for vulnerable student populations, we need a basic understanding of its genealogy and its evolution. This informed understanding will challenge us with a crucial question: If any of those century-old ideas and beliefs are no longer accepted, either because of contemporary research or our commitment to equal opportunity and antiracism, should we continue to use grading practices on which they were based?

The Twentieth Century Context

We begin our history at the end of the 1800s. Prior to that, in the first half of the nineteenth century, the family was primarily responsible for educating children, with schools serving a relatively small role. Relatively few children attended any formalized school—around half of white children ages fifteen to nineteen, and far fewer children of color who were either enslaved or, if they were free, still faced legal and extra-legal restrictions on their access—and the school year averaged only seventy-eight days (Snyder, 1993). In each school (which sometimes was simply a single room, the "one-room schoolhouse"), students of different ages learned side by side with age-appropriate curricula that often consisted of whatever books or other materials were available. The teacher herself—and she almost always was a woman—may or may not have received formal training to teach.[1] This model accommodated a nation organized around agricultural economies, independent proprietorship, and rural populations (Tyack, 1974).

In the early decades of the twentieth century, the United States experienced radical social, economic, political, and scientific changes that demanded changes to its schools:

1. *The rise of manufacturing*. While in the early 1800s, most people earned a living through agriculture or as craftsmen, by the turn of the twentieth century, American productivity had exploded and factories became the primary employers. In 1860, the United States lagged behind England, France, and Germany in its industrial output, but by 1900, it led the world and produced nearly as much value as those three countries combined (Tyack, 1974). Owners of factories needed workers, and they put pressure on school boards and city leaders to create schools that prepared their future

[1]If the teacher had any training, it was from enrollment in normal schools created to train high school graduates to teach (called "normal" schools because the schools were attempting to establish a norm or standardized model of teaching, borrowed from the French, *école normale*.

employees. There was also a cultural veneration for the power and productivity of factories, which persuaded policymakers to incorporate characteristics of industry—specialization, chain of command, timed routines, and efficiencies—into public institutions, including schools.

2. *Migration and immigration*. The lure of cities' manufacturing jobs and the modernized services (including water and sewage), along with a stronger railroad system that made travel easier, pulled people from their rural towns to the urban cities. While in 1820 there were only four U.S. cities with populations over 25,000 people, four decades later, thirty-five cities had populations of over 25,000, with nine cities of over 100,000 (Tyack, 1974). In addition, a massive wave of immigrants from Western Europe, and then Eastern Europe, came to the United States for jobs, and at the same time, cheap U.S. grain exports drove them out of employment in their home countries. By 1910, 40 percent of the entire U.S. population had foreign-born parents (Bowles & Gintis, 1976), and at around the same time, 58 percent of students had fathers who were born outside the United States, from over fifty countries (Tyack, 1974). The radical changes in the school-going population profoundly affected how we thought about schools.

3. *Progressive educators*. John Dewey and others envisioned that the realization of our still emerging democracy depended on an education that was "universal," integrated students from all backgrounds, provided opportunities to elevate one's social and economic position, and supported one's moral development. A less charitable perspective is that Progressives wanted schools to acculturate and assimilate the influx of immigrants so they were prepared to contribute to both democracy and capitalism. While many Progressives advocated for making school attendance compulsory and more standardized—a "common" school in which all students would be offered the same curriculum—others believed that differentiated education would address and accommodate the specialization of work in factories. In the end, although Dewey's vision of schools-as-democratic-engine provided overarching rhetoric about schools, it was often eclipsed by the vision of schools-as-training-ground. Bowles and Gintis (1976) write,

   In the end, the role of education as capitalist expansion and the integration of new workers into the wage-labor system came to dominate the potential role of schooling as the great equalizer and the instrument of full human development. (p. 181)

4. *Intelligence testing and categorization*. By the turn of the twentieth century, scientists had been exploring and theorizing about "natural intelligence"—the idea that one's mental ability was innate, immutable, and could be quantified by a range of assessments including those based on phrenology, the study of how a person's intellect and other characteristics are correlated

to the physical shape of the skull. The use of intelligence testing, stemming from Alfred Binet's research in the early 1900s, expanded dramatically in World War I when there was a need to quickly assign roles to the millions of enlisting servicemen. Scores on these tests soon became viewed as a reliable description of one's intellectual capacity, character, and disposition, and that provided seemingly scientific explanations and justification for racist beliefs. When Black people and immigrant groups from southeastern Europe and the Mediterranean scored lower, their scores were ascribed to weaknesses in intellectual capacity, nature, and upbringing rather than to the cultural biases of the tests or to the idea that those trends reflected gross social inequities associated with poverty or oppression. Plus, it was believed that intelligence, like other characteristics, occurs across a population with a normal distribution, termed a "bell curve" because of its shape. Low scores among people of color and immigrant groups, and the higher scores of white, wealthy Protestants, because of the design of the test, fit into a bell curve, tautologically justifying the validity of intelligence testing. Lower scores among immigrant groups and Black people were used both to affirm the idea of the United States as a meritocracy and to reinforce the legitimacy of the existing, inflexible hierarchy, now proven "scientifically."

5. *Behaviorism*. The first half of the twentieth century saw the popularity of behaviorism—the strand of psychology that argues that all human and animal behavior is the result of external stimuli, responses, learning histories, and reinforcements. It drew on Pavlov's findings from the 1890s that external stimuli could cause a reflexive effect: Dogs salivate when they see food, but if you introduce the stimulus of ringing a bell each time you show food, the dogs will be conditioned to salivate when you just ring the bell. John Watson built on Pavlov's ideas to argue that, similar to animals, humans are profoundly affected by their environment. B. F. Skinner took behaviorism one step further with his "Skinner box" experiments in the 1920s and 1930s, in which he taught rats to press a lever through two methods. He gave them food when they pressed the lever—"positive extrinsic motivation"—and alternatively would send electricity through the cage wires, causing pain to the rat until it pressed the lever, when he would turn off the electricity—"negative extrinsic reinforcement." He argued that one could increase or decrease a subject's voluntary behaviors through associated stimuli—"operant conditioning." This theory of learning—that animals, including humans, could be taught to act in certain ways through extrinsic reinforcement or consequences—became wildly popular.

Impact on Schools

These five trends powerfully influenced twentieth-century schools and their grading systems. The migration to cities (whether from rural areas or other countries) along with compulsory education laws in the first decades of the twentieth century

resulted in a huge increase in the student population. High school enrollment alone grew from approximately 203,000 in 1890 to 1.6 million in 1918 and created a need not only to increase the number of schools but also to change the number of students each school could accommodate. On average over this period, over one new high school was built every day, and each was built to serve hundreds, and even thousands, of students across multiple neighborhoods (Tyack, 1974).

Now that schools served many more students with a much wider diversity of backgrounds, languages, ethnicities, and incomes, there were two fundamental shifts in the purposes and design of schools. First, whereas schools had always been responsible for acculturating students into their community, the one-room schools had served a relatively homogenous group of students from families deeply rooted in the community with norms that were familiar; now, schools were expected to standardize and "Americanize" the diverse collection mass of immigrants, rural transplants, and the poor by preparing them with the discipline and habits that factories prized in its assembly-line laborers. In a document signed by seventy-seven college presidents and city and school superintendents of schools in 1874, the authors endorsed that schools should teach obedience and very specific skills:

> *Great stress is laid upon (1) punctuality, (2) regularity, (3) attention, and (4) silence, as habits necessary through life for successful combination with one's fellow-men in an industrial and commercial civilization. (Harris & Doty as cited in Tyack, 1974, p. 50)*

Behaviorism made this expectation possible; with the right combination of reinforcement and consequences, any student could learn to act in desired ways.

Second, charged with preparing students to meet the needs of the industrial and commercial world, schools could do so most efficiently if they matched each student with the appropriate curriculum based on the student's ability—the Progressives' vision of schools-as-training-ground. Equipping each student with the skills most appropriate to their intellectual ability would create the smoothest and most successful transition into the work world, and this would lead to economic success for the country. The director of the Bureau of Research and Guidance, in Oakland, California, for example, believed it was critically important to "find the natural ability of the pupil and place him where he belongs" (Dickson, 1922). If a student did not possess the intellectual capacity to succeed in a more rigorous academic track, then to not match that student with a vocational track would be a waste of school resources and would frustrate the child, perhaps leading to dropping out and depriving the commercial world of the student's contribution. To help schools efficiently place each student "where he belongs," a group of psychologists adapted the scales of the U.S. Army's IQ test, used to screen enrollees for officer training, to place students into different academic "tracks." Tracking students based on their intelligence test scores quickly became commonplace. In

1919, 400,000 copies of the "National Intelligence Test" were sold within the first six months on the market, and by 1920–1921, approximately two million children were tested (Tyack, 1974). By 1932, of the 150 American cities with populations over 100,000, 75 percent of them used IQ tests to assign students into schools' ability tracks. Detroit, for example, divided its students citywide into the top 20 percent, middle 60 percent, and bottom 20 percent of scorers and placed students of each group into different course sequences (Tyack, 1974).

Tracking students to situate them for specific roles in the economic hierarchy helped replicate the existing social and racial hierarchy and provide "scientific" justification for doing so. Schools assigned African Americans, immigrants, and lower-income student groups to lower tracks designed to teach them behaviors and skills that consigned them to reap fewer opportunities and a smaller share of the American Dream, and this inequity was normalized. Cubberley (1909), a prominent educator and scholar during the turn of the century and dean of the Stanford School of Education for nearly two decades, wrote that urban schools should "give up the exceedingly democratic idea that all are equal, and that our society is devoid of classes" (pp. 56–57). Some went even further. In 1924, Frank Freeman, a writer for *Educational Review*, wrote:

> *It is the business of the school to help the child to acquire such an attitude toward the inequalities of life, whether in accomplishment or in reward, that he may adjust himself to its conditions with the least possible friction. (as cited in Bowles & Gintis, 1976, p. 102)*

Dewey, who saw the institutional problems in society as being more about income inequality rather than race, recognized how deeply flawed this strategy was, not only for educators' reverence for testing but also because it undermined schools' democratizing function. He derisively asserted that:

> *Our mechanical, industrialized civilization is concerned with averages, with percents. . . . [W]e welcome a procedure which under the title of science sinks the individual in a numerical class; judges him with reference to capacity to fit into a limited number of vocations ranked according to present business standards; assigns him to a predestined niche and thereby does whatever education can do to perpetuate the present order. (Dewey, 1922/1983, p. 297)*

It's also important to keep in mind that schools' new commitment to evaluating students and sorting them occurred alongside a legal sorting of Black students who, in many parts of the United States, were constitutionally mandated to attend separate and unequal schools (see Gándara & Hopkins, 2010; Winzer, 2009).[2]

[2]While in this book it is neither possible nor my intent to describe the discriminatory mechanisms—both formal and informal—applied to multilingual learners and students with special needs, they were similarly sorted either into lower academic tracks or were viewed as ineligible or inappropriate for formal schooling.

It's easy to see how these ideas—schools as sorting and acculturating mechanisms in service to efficient and appropriate preparation for workforce employment—remain pervasive 100 years later. Tracking in our schools persists despite evidence of uneven pedagogical benefit and its discriminatory results: Students from low-income families; Black, Latinx, Indigenous, and other students of color; and those with special education needs are disproportionately placed in vocational and lower-track classes, and those classes have been consistently found to have lower academic expectations and engaging pedagogy (see Anderson & Oakes, 2014; Borman & Dowling, 2010; Lee et al., 1988; Oakes, 1990). In addition, the largest industries (currently, computer technology) constantly exert pressure on schools to provide appropriately trained employees. Schools continue to serve as assimilating and socializing agents, and though twenty-first century industries often demand more advanced skills than the assembly-line factory owners a century ago, in many classrooms, we continue to place a premium on punctuality, quiet attention, and following directions, the same behaviors desired of students over a century ago.

Grading in the Twentieth Century

With our understanding of how American schools in the early twentieth century reflected the zeitgeist of the country, let's look briefly at the evolution of grading and how it reflected and facilitated schooling.

As we mentioned earlier, prior to the turn of the century, before the large influx of families to urban centers and the rise of large schools to accommodate their children, the one-room school served few students and the teacher was a familiar member of the tight-knit community. It therefore should come as no surprise that communicating student progress looked very different than today. In most cases, the teacher would present oral reports or written narratives to families, perhaps during a visit to a student's home, describing performance in skills like penmanship, reading, or arithmetic (Guskey & Bailey, 2001). These reports helped determine areas for the teacher's further instruction for the student, readiness for apprenticeships, or eligibility for higher education (Craig, 2011).

Schneider and Hutt (2014) write that with compulsory education laws, schools became the "centre of a society increasingly dominated by complex bureaucratic institutions, including the school system itself" (p. 5). With larger schools and the emphasis on efficiency, schools had to develop more succinct and simplified descriptions of student progress. Schneider explains a significant shift:

> *Grading systems that had traditionally tended toward the local and the idiosyncratic, and which were designed for internal communication among teachers and families attached to a given school, became forms of external*

> *communication and organisation as well. Increasingly, reformers saw grades as tools for system-building rather than as pedagogical devices—a common language for communication about learning outcomes. (p. 5)*

No longer could educators use idiosyncratic narrative reporting—it was time consuming and too unstandardized. Instead, there was pressure to identify a simple system of communicating student achievement, not only for bureaucratic ease within the school for sorting purposes but also for external audiences—colleges or employers.

K–12 schools looked to higher education for guidance. Letter grades (A–F) had been in place in renowned European colleges and universities for centuries,[3] so American colleges and universities, informed by those models, created systems in the early 1900s to signify a student's achievement in a course relative to others in the course—called "norm-referenced grading"—and secondary schools began to use the letters well (Cronbach, 1975, as cited in Schneider & Hutt, 2014). Because, as the thinking went, intelligence is spread across a population with a normal distribution just like height or weight, then grades are more objective when they reflect that curve within any population. Schools therefore superimposed the normal distribution across a student group and labeled them by letter according to that distribution. By the mid-1900s, a majority of secondary schools used A–F grading and assigned grades according to the normal curve distribution, with the letter grades often demarcating segments of the 0–100 scale.

Throughout the twentieth century, schools and universities continued to use this same approach to grading, and although there have been critiques and some proposed alternatives, the Industrial Revolution's grading system remains relatively intact, with its founding beliefs embedded. Our twenty-first century schools may be dramatically different than schools of the twentieth century—the desks arranged for collaboration (rather than bolted to the floor), the laptops and iPads (rather than books), and the smart boards and PowerPoint presentations (rather than chalkboards)—yet still for each class, students receive a letter grade, translated from the 100-point percentage scale, that represents their performance. In many classrooms, those grades are assigned with the normal curve in mind, and these grades are used to sort students into different tracks and opportunities. In the next chapter, we hold up century-old grading practices, driven by century-old beliefs and interests, against our contemporary research and understanding: Could our best thinking about effective teaching and learning be thwarted by century-old grading?

[3]Cambridge University as early as the sixteenth century used a three-tier grading system with 25 percent of the grades at the top, 50 percent in the middle, and 25 percent at the bottom (Winstanley, 1935, as cited in Winter, 1993), a blunt ancestor of the bell curve.

Summary of Concepts

1. In the first half of the twentieth century, our country experienced radical social, economic, political, and scientific changes that included the rise of manufacturing, mass emigration from foreign countries and rural communities to cities, progressive educational theory, intelligence testing, and behaviorism. All these shifts influenced the transformation of American schools during this period.
2. Schools during this time were expected to assimilate large numbers of students into "American" culture, specifically to make them ready to be employed by factories. This meant that there was a priority placed on both teaching students certain behaviors suited for factory labor and to replicate the ethos of industry: efficiency and productivity.
3. While grading had previously existed as a teacher's narrative of student progress, twentieth century schools adopted single letter grading (A, B, C, D, F) and the use of the bell curve to more efficiently describe and communicate student performance and to sort students easily.

Questions to Consider

1. How do schools in the first half of the twenty-first century—their design, their purpose, their students—compare to schools in the first half of the twentieth century?
2. How do you see the ideas and beliefs of the early twentieth century manifesting themselves through your school's communication, curriculum, instruction, policies, and grading?
3. In the last century, how has society changed, and how has this affected what we want students to learn in school—the skills, knowledge, and ways of working? How has our understanding of the teaching profession changed? How would we adjust grading to reflect these changes?