

This article was downloaded by: [University Of Maryland]

On: 19 November 2013, At: 09:17

Publisher: Routledge

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



Journal of Curriculum Studies

Publication details, including instructions for authors and subscription information:

<http://www.tandfonline.com/loi/tcus20>

Making the grade: a history of the A-F marking scheme

Jack Schneider & Ethan Hutt

Published online: 16 May 2013.

To cite this article: Jack Schneider & Ethan Hutt , Journal of Curriculum Studies (2013): Making the grade: a history of the A-F marking scheme, Journal of Curriculum Studies, DOI: 10.1080/00220272.2013.790480

To link to this article: <http://dx.doi.org/10.1080/00220272.2013.790480>

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis makes every effort to ensure the accuracy of all the information (the "Content") contained in the publications on our platform. However, Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Any opinions and views expressed in this publication are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor and Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden. Terms & Conditions of access and use can be found at <http://www.tandfonline.com/page/terms-and-conditions>

Making the grade: a history of the A–F marking scheme

JACK SCHNEIDER and ETHAN HUTT

This article provides a historical interpretation of one of the defining features of modern schooling: grades. As a central element of schools, grades—their origins, uses and evolution—provide a window into the tensions at the heart of building a national public school system in the United States. We argue that grades began as an intimate communication tool among teachers, parents, and students used largely to inform and instruct. But as reformers worked to develop a national school system in the late nineteenth century, they saw grades as useful tools in an organizational rather than pedagogical enterprise—tools that would facilitate movement, communication and coordination. Reformers placed a premium on readily interpretable and necessarily abstract grading systems. This shift in the importance of grades as an external rather than internal communication device required a concurrent shift in the meaning of grades—the meaning and nuance of the local context was traded for the uniformity and fungibility of more portable forms.

Keywords: grades; report cards; student evaluation; grading

Grading is one of the most fundamental facets of American education. In hundreds of thousands of modern US classrooms, grading is a well-accepted part of schooling. It is as natural as the use of textbooks, or the arrangement of students in desks or the presence of a teacher in the room; it is a part of what Tyack and Cuban (1995) have called the ‘grammar of schooling’. From the elementary grades up through advanced graduate programmes, instructors spend hours each week correcting papers and exams, students wring their hands over grades and steal glances at each other’s scores and parents express various levels of anxiety about the marks their children are earning. Grades have become such an important feature of adolescence that we have even invented short-hand references like the grade point average so that a student’s academic record may be expressed with the seeming precision of a single number and judged at a glance. And unlike so much of the ephemera of adolescence, grades have lasting and profound consequences: once earned, they serve as a key determinant of future success—a mechanism through which schools, universities, and employers judge the individual’s academic achievement.

Of course, grading is not without its critics. Since the inception of grades detractors have noted the many problems and perverse incentives they create. Some have argued that grading does psychological harm to

Jack Schneider is an assistant professor of Education at the College of the Holy Cross, Worcester, MA, USA; e-mail: jschneid@holycross.edu. He is the author of *Excellence for All: How a New Breed of Reformers Is Transforming America’s Public Schools*.

Ethan Hutt is an assistant professor of Education at the University of Maryland, College Park, MD, USA. He is currently completing a book on the history of minimum educational standards.

children (e.g. Kirschenbaum *et al.* 1971; Kohn 1999). Others have made the case that grades are too simple a measure for gauging school success (e.g. Tough 2011; Wilson 2009: 58–62). And still others have suggested the extrinsic motivation of grading works against the fundamental purposes of education (e.g. Beck *et al.* 1991: 35–37; Butler 1988: 1–14; Milton *et al.* 1986; Pulfrey *et al.* 2011: 683–700). Still, those opponents and their critiques have never occupied more than the margins of the discussion over grading and they have done little to dislodge formal marking systems.

Grading remains a central feature of nearly every student's school experience. As such, it can be easy to perceive grades as both fixed and inevitable—without origin or evolution. And the effect of this is that despite their limitations, grades are often accepted quite uncritically by all parties involved.

Yet grades have not always been a part of education in the United States, they have not always looked the same, have not always served the same purpose, and have not always had the same impact. A discussion of these matters may, then, promote more critical understanding of the current place of grades. It may give us a way of making sense of their necessity, purpose, utility, and potential.

Given the promise of such an analysis, it is somewhat surprising that historians have paid little attention to the topic. Consequently, even the most basic questions concerning the origin and diffusion of grades and grading systems—where they came from, what they were supposed to achieve, and why they took the form they did—remain largely unanswered. This article aims to remedy that by presenting a critical and accurate accounting of the past.

In order to do this, we locate grading systems in the broader history of education in the United States because, as we will show, the history of grades reflects the interplay between the work of the classroom and society at large. As the little red school house of the frontier became increasingly linked to the great universities of the east and the labour markets of the continent, grades were required to serve a growing list of functions. Tracing this development, we situate grading schemes as a key technology of educational bureaucratization, a primary means of quantification, and the principal mechanism for sorting students. In short, this article frames grading as a crucial expression of the modernist impulse.

Early American grading systems owed much to the European model—focusing on constant competition, the awarding of prizes and rank order competition—and were largely used for pedagogical purposes. The introduction of mass compulsory schooling, however, changed things dramatically. Mass schooling placed the school at the centre of a society increasingly dominated by complex bureaucratic institutions, including the school system itself. Consequently, grading systems that had traditionally tended towards 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 organization as well. Increasingly, reformers saw grades as tools for

system-building rather than as pedagogical devices—a common language for communication about learning outcomes.

If grades were to communicate beyond the school site, marking systems had to be made more ‘legible’, more universal and more standardized.¹ Driven by policy elites and eager administrators, grading systems expanded, reproduced and evolved. Without a central authority to mandate standardization or facilitate communication, they advanced in fits and starts varying across the regions and levels of schooling. Yet by the turn of the twentieth century, teachers, administrators, parents, college admissions officers and employers were turning to grades for basic information about academic aptitude and accomplishment. This is not to say that all parties saw grades as an unqualified good; in fact many educators expressed concern about the consequences associated with grades, particularly with regard to curriculum and instruction. Nevertheless, reform-minded policymakers continued to pursue the system-building aims of objectivity and uniformity well into the twentieth century.

This article offers an examination of the history of one of the most well-accepted elements in American schools, seeking to understand how and why an ‘A’ became an ‘A’ (and an ‘F’ became an ‘F’). In doing so, it considers grading systems as key instruments in two different processes—that of internal communication oriented towards pedagogical concerns; and that of external communication oriented towards system-building. Ultimately, in tracing the use of grades from the origins of formal education in the United States, this work tells a story about a core and seemingly inescapable tension in modern schooling: between what promotes learning and what enables a massive system to function.

Europe and the early American republic

In 1785, Yale president Ezra Stiles, after examining 58 seniors recorded in his diary that among the students there had been ‘Twenty Optimi, sixteen second Optimi, twelve Inferiores (Boni), ten Periores’ (Stiles 1901). Scores, as doled out by Stiles, were determined by the perceived learnedness of response. Equally, however, they were decided by a student’s ability to demonstrate knowledge publicly (Bristed 1874). And though this system of public examinations did not last long, it did reveal the extent to which American educators—at least initially—looked to their European counterparts for guidance.

Specifically, it appears that educators like Stiles were mimicking a classification scheme best exemplified by the Cambridge Mathematical Tripos examination. The Tripos was established in the first decades of the eighteenth century, and by 1790 had taken written form. Students competing in the Tripos, which resembled a multi-day academic tournament, were ranked prior to the beginning of the examination and were ‘re-seeded’ after each day of testing (Searby 1997: 150–170). As students moved up in the brackets, they were challenged by increasingly difficult questions, as well as by increasingly tougher opponents—a fact that made these examinations fiercely competitive.

The rewards for performing well in these competitions were, perhaps, worth the considerable effort required to obtain them. Like titled lords, high scorers on the Tripos carried their distinctions forward with them and some stood to gain considerable financial awards. The highest scorer in the Tripos, for instance, would receive a portion of the university's endowment for the rest of his life (Searby 1997: 95). Low scorers, for their part, would remain marked by their performance, as well. An 1850 biographical sketch of Eton graduate Richard Porson, for instance, recalled that in 1777, Porson entered Trinity College at Cambridge and, because he 'paid but little attention to mathematics ... only took a Senior Optime's degree' (Creasy 1850).

American educators, however, did not merely mimic the models of highly visible institutions like Cambridge. While they shared the European aim of motivating students, the idea of fixed titles struck many as anti-democratic. At Yale, Stiles's categories for student ranking were soon 'replaced by a system of honors 'appointments' running from Orations at the top to Dissertations, Disputes and Colloquies; and these groups were supplemented by a system of grades running from 4 to 0 (Pierson 1983: 310). In this numerical system, originally applied only to upperclassmen for work on oral examinations, a score of 2 earned students a passing mark. Yale soon began using quarter points to further sub-divide the system and extended it to younger students in 1819 and 1820. By 1837, 'performances in individual courses were graded' and, like grades for work in oral exams, were dutifully recorded in the 'Book of Averages' (Pierson 1983: 310). These grades, however, were often kept secret from students—an effort to minimize the day to day competition among them (Bagg 1872).

Yale was not the only school experimenting with ranking and grading schemes. William and Mary faculty reports in 1817, for instance, indicated that four categories were established to distinguish students: the 'first in their respective classes'; those who were 'orderly, correct and attentive'; those who made 'very little improvement'; and those who learned 'little or nothing ... on account of excessive idleness'. As the labelling of the second category—'orderly, correct and attentive'—makes clear, this grading system took into account more than just academic achievements (William and Mary College 1817: 5). And the college shared the ordering of students into these particular categories, as well as notes on the propriety of student conduct, with parents and guardians (William and Mary College 1817: 15). Other colleges also issued grades for non-academic criteria. Students of both Yale and Harvard, for instance, found themselves graded on whether they attended chapel or showed up to class (Peabody 1888).

Grammar schools and high schools had also begun issuing grades in the early- to mid-nineteenth century. In the appendix to his 1824 book, for instance, British teacher John Shoveller included 'an example of the manner of calculating a week's work' in which he graded student work by subject as 'optime, bene, pessime, male', using superlatives and diminutives like 'sen' and 'jun' to further differentiate quality. At a school in Glasgow, students were examined eight times each year by a Committee

of the Town Council, Clergy, and Professors. Their scores were ‘carefully marked on all these occasions, and their *average rank* in the Class [was] ... calculated from these examinations’ (Russell 1826a: 571). Given the high levels of migration from Britain to the United States in this early period, it should come as no surprise that many American schoolmasters implemented similar kinds of schemes. And as was the case in higher education, grammar and high school students were graded according to both academic and non-academic criteria (Labaree 1988).

The European schools whose grading practices received the most attention from American educators were the Prussian schools. Horace Mann and his fellow reformers followed the organizational innovations of the Prussian schools with great interest, and saw them as both pedagogical and organizational models (Mann 1845). Particularly influential among this group of reformers was an 1837 report by Calvin Stowe in which he noted that Prussian schools organized children and the curriculum in terms in a series of stepped grades that allowed students to move along at their own pace, while increasing the overall efficiency of the system (Stowe 1838).

Prussia’s graded system appealed to Mann and his colleagues as a significant improvement over the way American schools were organized. Many schools, particularly in rural areas, had no formal record-keeping mechanisms—in part because students attended irregularly, in part because students of different ages were grouped together, and in part because there were no common texts with which to work (Kaestle 1983). Many schools in urban areas, by contrast, were organized after the Lancasterian, ‘monitorial’ model. Under the Lancasterian model, students were frequently subject to examinations and direct ranking against their classmates. Each day, students finished their assignments and were literally repositioned in the classroom—the top students moving to the front of the class and the less capable students moving to the back (Kaestle 1983).

Advocates of the Lancasterian system cited its cost efficiency. But they spoke in far more glowing terms about the frequency of competition inherent in it and the fluidity of its ranking—features praised as both commendable and appropriately distinctive to the United States. A description of Boston ‘monitorial’ school from 1826, for instance, noted that the highest performers in spelling lessons were appointed as monitors, followed by ‘the highest class’ and rank ordered down ‘to the lowest’ (‘Boston Monitorial School’ 1826: 34). Such communication to students about their relative standing was, at least by some, thought to motivate students. As one author noted, ‘those who have lost their rank’ to their classmates would work to recover it, ‘especially as *their* failure may have been a means of [a classmate’s] gain, without any merit on his part’ (Russell 1826b: 561).

The pedagogical concern with student motivation was common in the period. Gideon F. Thayer, former principal of Chauncy-Hall School in Boston, echoed such a sentiment in his ‘Letters to a young teacher’ in 1856:

Tell the scholars that, as soon as you shall have become acquainted with them, you intend to establish a 'merit roll', and that you cherish the hope that all, or with few exceptions, will have a claim to the front rank ... By thus showing them that they all have it in their power to distinguish themselves, whatever their scholarship, you may be able to enlist a large number of allies in your work, which will hence go on all the more prosperously, because adopted cheerfully, or from choice. (661)

Such sentiments were not limited to K-12 schools. In 1831, the *Annual Report of the President of Harvard University* stated that 'the best assurance for the continued and unremitted attention of students to their exercises ... [is] the certainty that at every recitation each individual will be examined; and that the estimate of scholastic rank must depend, not upon occasional brilliant success, but upon the steady, uniform, and satisfactory performance of each exercise' (Harvard University 1831: 285). In short, reformers saw in such competitions the key to student motivation.

Still, policy-makers worried about the message that this constant competition sent to students and they expressed misgivings about the nature of school reward structures. The constant ranking and re-ranking placed emphasis on the immediacy of competition rather than on intellectual and moral development. As Horace Mann wrote in his ninth annual report, 'if superior rank at recitation be the object, then, as soon as that superiority is obtained, the spring of desire and of effort for that occasion relaxes' (Mann 1846: 504). Pupils, he imagined, might focus so intensely on the outcomes of examination 'as to incur moral hazards and delinquencies' (505). Moreover, the shifting nature of those ranking systems made it difficult to use them as a means of communicating beyond an individual school.

The solution to this problem, according to Mann, was not only to transform schools from one giant competition into a series of graded steps, but also to substitute the public quizzes and frequent re-ranking for written examinations and a series of monthly report cards. Thus, as Mann saw it, a student might come to understand that 'his mind will be submitted for inspection not only on its bright side, but on all the sides; and that it will be useless for him to expect to shine on that occasion with only a radiant beam of light thrown across it here and there, while wide intervals of darkness lie between' (Mann 1846: 508). Reformers likened these report cards to merchant ledgers, which emphasized the accumulation of success over time and provided a running account of a student's academic success ('School records, 1864').

Just as importantly, this less overt form of grading would reduce the general level of competition among students within a school—a move that complimented the broader organizational move to age-grading. No longer in direct competition with each other to get to the 'head of the class', students were free to be more collegial and less competitive. The result, reformers believed, would return the focus of education to the intrinsic value of learning rather than the extrinsic motivation of academic acclaim (Barnard 1854; Shearer 1899). Reformers also believed that monthly report cards would allow teachers to keep parents informed about their children's achievements. An 1835 issue of the *American Annals of Education*, for example, recommended a 'weekly report to the parents, exhibit-

ing in a compendious manner, the punctuality, deportment, and comparative merit of the pupil, in his recitations' (Woodbridge 1835: 525).

Taken together, these developments show how much American grading systems diverged from their early European origins. As the system of common schools took root in America, reformers recognized the need for grades to act as important internal organizational devices—to maintain student motivation while minimizing competition and emphasizing the accretion of knowledge.

But though reformers were coming to a general consensus about the purpose of grades, they had yet to standardize the practices themselves. That was a task that would take on increasing importance as both the public education system and society, as a whole, grew more complex in the last decades of the nineteenth century.

Building a national system

By the end of the Civil War, grades were a relatively normal part of primary, secondary, and higher education in the United States. Despite this normalcy, they were still determined in a highly disparate fashion, they were represented in various ways—lettered systems, percentage systems and other numerical systems—and schools modified and refashioned grading systems with considerable frequency (Smallwood 1935). At Yale, in the 1870s, for instance, grading was briefly conducted on a 200–400 scale, whereas at Harvard, students were classified in six divisions based on a 100% basis (Pierson 1983, 312; Smallwood 1935, 50–52). At some schools, behaviour was calculated into grades whereas at others, it was not (Harvard University 1869: 47). And, at some schools, grades were given only at the end of the school year, while at others grades were issued in periodical 'report cards'.

But times were changing. Educational enrollments were expanding rapidly, particularly in K-12 education—the product of compulsory schooling enforcement, child labor laws and the growing use of education as a tool for social mobility (Labaree 1988; Steffes 2012; Tyack 1977). Thus, between 1870 and 1910, K-12 enrollments almost tripled in size and as schools grew to absorb these students, administrators found themselves at the helms of massive institutions (Goldin 2006). In the face of this rapid expansion and depersonalization of schooling, administrators refashioned themselves as professional managers whose job was to manage burgeoning systems in the most-efficient way possible (Hansot and Tyack 1986).

Reformers saw grades as a means of creating modern systems for a modern world. In increasingly massive urban systems, teachers could no longer give detailed accounts of every student's abilities. Yet this was essential for other parts of the system to work. If students were to move from one grammar school to another, for instance, or from grammar school to high school or high school to college, they would need to be tracked in some systematic way.

Students certainly had moved in previous decades from school to school or from one level to another. But in the late nineteenth century,

they did so with increasing frequency (Snyder 1993). Traditionally, for instance, college entrance examinations were conducted by individual professors at individual high schools. The lack of standardization, however, left the results of these tests open to claims that the administration and marking of these examinations was too idiosyncratic to be fair, to say nothing of the tremendous effort required to carry it out (Schudson 1972; Wechsler 1977). Colleges moved to accredit high schools based largely on their curricula—a workable solution as long as the market for matriculation remained regional (Schudson 1972; Wechsler 1977). But the proliferation of high schools and the increased mobility of students required greater standardization. And though this problem would ultimately be solved as much by the creation of the College Entrance Examination Board (CEEB) as by standardization of grading practices, the creation of the CEEB nevertheless speaks to the increasing complexity of the education system and the need to develop more efficient, standardized, and coherent coordinating mechanisms for the United States's burgeoning school system.

These developments in education paralleled similar moves towards developing standardized grading systems in a host of other industries. Cronon (1992) and others have documented the tremendous amount of work that was required to standardize products ranging from corn to lumber to cattle in a way that made it possible to create national markets.² As these other growing industries were revealing, markets required their own pricing and rating systems in order to function at a larger scale. Hill (1990: 4), for instance, has illustrated the degree to which the volumes traded on a growing grain market made personal inspection by the buyer more and more difficult. Consequently, 'the communication and information provided by uniform grades and standards were required for an efficient marketing system to develop' (Hill 1990: 1). This was the reality of systems that could no longer rely on intimate relationships or knowledge between buyer and seller. Mass processing, whether of grain or academic achievement, required standardization.

Grades, then, would create for the rapidly expanding American educational system a unified and scalable mechanism for measurement and communication (Carruthers and Stinchcombe 1999: 353–382; Epstein and Timmermans 2010: 69–89). As Fischel (2009) has argued, the considerable mobility of the Americans during the nineteenth century placed pressure on school districts to standardize their practices—such as school organization and the synchronization of school calendars—despite the decentralized character of American education. Standardized grading practices were similarly implicated in this mobility story and the creation of increasingly national labour markets and the tighter vertical integration of school systems. In other western nations, the timing and particulars were different, but the idea was the same: the system of ranking stripped of its prizes, and ostensibly made objective through the use of a quantifiable marking scale, was a prerequisite for a modern school system—a system characterized not only by its universality, but also by its scalability (Compère 1985: 83; Van Herwerden 1947: 41).

By the 1870s, the schools were taking more aggressive actions to standardize grades. Like rankings of students, grades could indicate to

students something about their performance and serve as a motivational tool, all without creating rivalry and excessive competition between students. Unlike ordered rankings, however, which only communicated a student's relative standing within his or her class, grades also promised to serve as an external communication device—to admissions boards, employers and others. In order for either form of communication to be successful, though, grades would have to mean something similar across classrooms, and ideally across schools.

One manifestation of the effort to ensure greater uniformity was the move from evaluating student work over the course of a number of years to grading individual courses worth so much 'credit'. After the Civil War, for instance, the University of Michigan began defining degree requirements in terms of 'full courses', which consisted of five exercises a week for a semester (Hinsdale 1906: 81). And by 1892, Michigan was listing credit hours after every course in the catalogue (Gerhard 1955: 654–668).

Such standardized data sets allowed schools to measure their students in new ways. One way was the measuring of 'passes' and failures. According to the 1875–1876 *Annual Report of the President of Harvard University*, for instance, 133 students from that academic year had 'passed in all their work'; 11 'failed, but can make up'; and two 'failed, and must repeat the year' (Harvard University 1876: 55). A decade later, grades were used to determine the relative merit of students receiving tuition aid against their peers paying full price. Sixty-five percent of students receiving aid, it turned out, earned an A or B average (Harvard University 1889).

Some worried that while grading would achieve formal standardization, it would come at the price of precision. While once Harvard and other schools had taken seriously the task of rank ordering their students, Eliot recognized that this was no longer feasible, let alone possible. The elective system installed by Eliot meant that students were no longer taking identical courses—a move later paralleled at the high school level with the tracking of students—making it hard to produce the traditional list of students' ranks. And though Harvard experimented for a while with a 'Book of Comparative Merit', this proved untenable.

Eliot's solution, in the late 1870s, was to create a new form of honor system that divided students into three broad categories: *summa cum laude*, *magna cum laude*, and *cum laude* (Harvard University 1878). Like the shift between some high schools from class ranks to grades, Eliot's system—borrowed from European universities—used categorized honors to convey information about student achievement without the burden of striving for an impossible kind of precision (Bache 1855: 27). It also continued the trend of attempting to reduce the direct competition between members of a class, which was increasingly viewed as uncouth. As one commentator put it, 'Futile and somewhat anti-moral is the plan proposed of trying to improve scholarship by persuading students to compete for class rank'. Taking on the popular analogy of sports and academic competition, he continued:

We are told that ... 'there is a close analogy between outdoor sports and those indoor studies which are pursued for intellectual development, espe-

cially in regard to the question of stimulum by competition'. As a matter of fact, men pull together in a boat for the glory of their college; the man who plays for his own oar or hand is not esteemed there or elsewhere. (Cattell 1910: 369–383)

Rather than emphasize *individual* glory, then, Eliot and others believed that grades should indicate and distinguish classes of men.

In K-12 schools, parents welcomed greater uniformity in grading as a way of receiving regular and concise information about their children's abilities and achievements. As Francis Parker observed in 1902, 'parents in general measure school progress of their children by per cents on monthly report cards, by text-books finished, examinations passed, promotions gained' (762). Those districts and states without report cards, then, often turned to this new technology for communicating with parents. A 1912 article from *The Elementary School Teacher* reported that the state superintendent of instruction in Missouri, in coordination with the state's rural school inspector, had devised a plan to foster cooperation between school and home through the use of a report card. Interestingly—and as a signal of the lack of standardization at the turn of the century, the Missouri report card did not just communicate information from the school to parents; rather, it was also designed to allow parents to communicate to the school something about the work being done at home by pupils ('Educational News and Editorial Comment' 1912: 61).

By the dawn of World War I, most schools still pursued their own approaches with regard to grades. Highly visible schools, particularly among colleges and universities, often served as models. Yet there was little uniformity across the educational system and there was little coordination at the district or state level. Grades varied in terms of letters, percentages, or raw numbers, with generally even distributions among all. Half of high schools included a 'deportment' category, and half of elementary schools included an 'effort' category in grading. And there was no uniformity with regard to how often grades were issued (Ashbaugh and Chapman 1925: 291).

What had emerged by this period was the notion of all classes in a school issuing grades to students, and those grades amounting to an indicator of student ability and achievement. And as that notion became more accepted, grades began to communicate more and more powerfully outside of the school. As they did, however, they would stir up questions among those working *inside* of schools about just how well they were serving their original purpose of aiding learning.

Standardization and the bell curve

System builders in the early twentieth century continued much of the work of their predecessors—creating and refining grading systems, advocating for greater uniformity, sharing model practices, and developing new mechanisms for grading students. Consequently, more schools in the early 1900s began giving grades for individual courses, compiling records

of those grades, and adjusting their systems to square with the practices of other schools, districts, and states.

But though educational reformers continued to make progress towards their goal of developing grading systems, new questions began to arise about what was actually being measured. As one commentator noted, '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'. Despite this newly achieved uniformity, he continued, 'the problem now presented is that of establishing a method whereby grades assigned by one teacher can be intelligently compared with those assigned by another, and all brought to a common standard' (Weld 1917). In other words, it was no longer enough to have the same grading system. The next task was to demonstrate that grades everywhere had the same meaning.

This was not a problem unique to education. In grading commodities, for instance, the major obstacle to creating uniformity was 'the lack of objective, standard measures and instrumentation' (Hill 1990: 19). As one observer of the wheat industry noted in 1908: 'gradations are continuous, and if lines are drawn to mark the limits of the grades, it is difficult to determine the grades in cases close to the lines ... [and] their interpretation has been left largely with the grain inspectors' (Dondlinger 1908: 221).

In education, however, this concern was amplified, given the increasing influence that grades had on students' futures. As Isador Finkelstein wrote in 1912,

the evidence is clear that marks constitute a very real and a very strong inducement to work, that they are accepted as real and fairly exact measurements of ability or of performance. Moreover, they not infrequently are determiners of the student's career. They constitute the primary basis for election to honorary societies, for the award of various academic honors, for advancement from class to class, for graduation, and may even determine in some measure the student's career after leaving the institution in which they have been assigned (Finkelstein 1913: 6).

As one observer of the University of Missouri wrote in 1911: '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' (Meyer 1911: 661). And, as a third wrote: 'Diplomas are hallmarks of excellence like the chemical manufacturer's 'C.P.' guaranty ... [and must] be kept according to some uniform system intelligible to the people most vitally interested—the public' (Campbell 1921: 511). Objectivity, then would be elusive as long as grades were ultimately left to the discretion of each inspector.

Some, of course, favoured an end to modern grading. In 1918, Thorsten Veblen argued that the 'system of academic grading and credit ... resistlessly bends more and more of current instruction to its mechanical tests and progressively sterilizes all personal initiative and ambition that comes within its sweep' (128). In a 1932 article from *The Elementary School Journal*, the Assistant Superintendent of Salt Lake City Schools

reported on an experimental report card that in most cases contained only entries on attendance and punctuality (Worlton 1932: 179). An author writing in the *Junior-Senior High School Clearing House* observed that the report card, 'which was once considered to be of almost divine origin, has been challenged to mortal combat'. It was, as she noted, a 'challenge from a group of young crusaders who have chosen to be known as the Intrinsic Clan against the entire family whose surname is Extrinsic' (Sumner 1935: 340). Yet another author wrote: 'Following the influence of John Dewey, many educators have been trying to get activities in education to be motivated by intrinsic interest of the pupil. In this attempt they have attacked the giving of marks on the basis that the pupils tended to work because they desired high marks rather than because of their interest in the subject' (Segel 1936: 34).

But grades were not going to be dropped, however difficult the search for objectivity. They were too useful as an extrinsic motivation, even if such motivation was seen as having serious consequences. Arthur L. Campbell wrote in *The School Review* in 1921: 'that our marking systems of today are fraught with innumerable weaknesses and inconsistencies, their most loyal adherents cannot deny; on the other hand, that they do serve as a spur to the laggard, even their most outspoken opponents must admit' (511).

Grades had also become instrumental to the functioning of an increasingly complex national educational system. As such, those with a vested interest in the system, and not merely system-building reformers, had a real stake in maintaining established grading schemes. 'A majority of parents, and too many teachers', V.L. Beggs wrote in *The Elementary School Journal* in 1936, 'conclude that the school's most important contribution to the child's education is recorded on the card' (107). As a report on college enrollments noted, many schools 'which formerly admitted by examination' had moved to accept only 'students who rank high in the secondary school'—a tendency 'influenced by a number of studies which show a high relation between high-school rank and success in college' (Gladfelter 1937: 742).

The growth of grading systems in the early decades of the century was also the product of what historians have dubbed the social efficiency movement in American schools. Concerned that students should be taught what they would actually go on to use, leaders like Franklin Bobbitt and David Snedden advocated for a curriculum tailored to each student's future destiny. As Bobbitt put it, the aim was to 'educate the individual according to his capabilities' (1912: 269). Grades, then, would aid educators in gauging student ability, producing seemingly objective sets of academic records. As such, educational reformers concerned with this aim of social efficiency, as well as with 'scientific management principles', worked to promote the further entrenchment of grading schemes in K-12 schools (Kliebard 1987; Krug 1969).

At the same time, the mental testing movement spurred educators to further standardize K-12 grading practices in an effort to demonstrate that schools, like businesses and the military, were sorting people through precise measurement. After all, if intelligence could be measured through

examination and then conveyed through a single figure like IQ, then measuring student's performance on school exams or in discrete academic courses should also be possible (Sokal 1987; Wright 2005). Influenced by such thinking, policy-makers and school leaders worked to align their practices with contemporary thought about educational measurement. Primarily, this manifested in efforts to create grading schemes as seemingly 'objective' as mental tests were perceived to be, including particular aspects like distribution across a normal curve (Cronbach 1975). As Finkelstein wrote in 1913, 'native ability, from all the evidence at our command, behaves like any other biological trait'. As such, he reasoned, 'in any population its distribution is that known as the curve of error, the probability curve, or Gauss's curve' (11). Given this, school grades should be devised accordingly—'based upon the orientation of all students around a central group whose accomplishment is construed to be average or medium' (18).

Not all educators, of course, were influenced by the mental testing movement or convinced about the superiority of normal distributions. After all, the teacher in the classroom grading the work of a student was faced with a very different educational task than was the psychologist attempting to measure native intelligence. While measuring a child's 'innate' intelligence might be sufficient for sorting a student into a particular school track, grades were intended to reflect a level of achievement and educational attainment distinct from native intelligence. It was communicating this level of attainment and general classroom performance that teachers were concerned with in assigning grades. And general recognition of these distinct purposes is evidenced by the large number of studies that examined the correlations among measured intelligence, personality traits and school grades (e.g. Ames 1943; Bolton 1947).

The mental testing movement's inclination towards objective measurement and the desire to use grades as another variable in their efforts to develop more discerning and predictive tests caused many educators to bristle. In response to pressure to adjust grading practices, for instance, one teacher complained that 'it is time for the science of education ... to make [the normal curve] merely one of a class of teacher curves, all different and having function of significance. In fact the usefulness of the teacher curve ... has been scandalously overlooked' (Rutt 1943: 124). Such objections reflected a disconnect between policy-makers focused on system-building and classroom teachers charged with educating.

Though teachers could not deny that grades functioned as a sorting mechanism, they frequently rejected the notion that they had to issue them according to the parameters of a fixed distribution. Still, many acceded to the demands of policy-makers and administrators because of the inherent threat to teacher autonomy posed by the mental testing movement. After all, if educators wanted to maintain their claim on professionalism, they needed to demonstrate that their evaluations of student work were accurate and reliable. As such, while some teachers wrote in opposition to an increasingly rigid grading scheme, many at least outwardly accepted it and sought to present their grading practices as objective and scientific.

Reformers focused primarily on the issues of objectivity and uniformity saw much work for themselves in American schools (Dustin 1926: 28). A body of research developed in the 1930s and 1940s, for instance, indicated that ‘teachers, on the basis of their judgment and such tests as they would of themselves construct, cannot reliably mark pupils on the basis of percentages’ (Segel 1936: 34). Another commentator, in 1936, wrote in the *Review of Educational Research* that ‘variation among high schools of the scholastic aptitude of their graduates and the variations in the standards of marking are so great that any index based on school marks is subject to gross misinterpretation’ (Bixler 1936: 169). Some commentators, noting the disconnect between student achievement as measured by standardized tests and as reflected in teacher grades, would go so far as to wonder whether the grades were ‘fact or fancy’ (Hadley 1954).

Yet, while this motivated some to continue the search for objectivity, it drove others to experiment with qualitative rather than quantitative feedback. One author writing in the *Elementary School Journal*, for instance, suggested a ‘diagnostic letter’ in place of grades (Beggs 1936: 112). An experiment in Pasadena, CA involved students and teachers working together to evaluate the student’s work (Beggs 1936: 172). Another experiment in Newton, MA, centred on letters written to parents twice a year and accompanied by questionnaires (Beggs 1936: 172).

A teacher at Westwood High in Los Angeles reported that at his school ‘the suggestion of substituting a letter to the home in the place of a formal mark was considered at some length’. Two teachers designed evaluative outlines for letters that would replace report cards in their classes. Yet they recognized that any standard form, even narrative letters, would eventually ‘become as meaningless and as stereotyped as the subject marks they replaced’ (Geyer 1938: 531). Grades would need to be entirely eliminated if they were to avoid what opponents derided as ‘false standards of value among our pupils’ (Geyer 1938: 530).

But reformers placed increasing emphasis on legibility beyond the school and a mass of information in the form of a diagnostic letter would not allow for that. As one author wrote in 1938, ‘modern records’ were not merely for conveying information to parents; rather, they were ‘functional’ for a broad range of purposes (Wofford 1938: 185). And, as another wrote in a 1939 issue of *The Elementary School Journal*:

most of the report cards issued recently appear too heavily burdened with long check lists covering the various items that go to make up personality. Many of these are so complex and detailed that the teacher cannot be expected to rate them with any degree of reliability, nor can the average parent be expected to understand the meanings of all the marks. There is need here for simplification which will make for greater reliability on the teacher’s part and better understanding on the part of the home. (Kvaraceaus 1939: 747–750)

In short, by the end of the 1930s, schools were responding to pressure for legibility. They were simplifying, working to standardize, and struggling to implement some kind of internal uniformity. And externally, parents,

admissions officers, and employers looked to grades for information as well.

Although the A–F grading system was still not standard by the 1940s, it had emerged by that point as the dominant grading scheme, along with two other systems that would eventually be fused together with it: the 4.0 scale and the 100 percent system. This move was slow, of course—the product of a decentralized system with few formal coordination mechanisms. Yale, for instance, went through four different grading systems between 1967 and 1981, moving from a numeric system to an Honors/High Pass system to the A–F system and, finally, to an A–F system that also included pluses and minuses with letter grades allowing for additional precision and sorting (Yale Ad Hoc Committee on Grading 2013). But uniformity, in K–12 schools as well as in higher education, was slowly emerging in the United States. And as educators would soon find out, there was no turning back.

Mutations and resistance

By the mid-twentieth century, grades in American schools had become largely standardized. Most K–12 schools and colleges issued grades of A, B, C, D and F to students, and those grades generally aligned with numerical values—an A reflecting work between 90 and 100, for instance, and a B reflecting work between 80 and 89. By the 1960s, the A–F system was being called ‘traditional’ and the practice of translating letter grades into numbers—A = 4, B = 3, C = 2, D = 1, F = 0—was being called ‘familiar’ (Burke 1968: 12; Chansky 1964: 95). According to a National Education Association survey (1974), letter grades were in use in over 80% of schools by 1971.

As this system took hold, other changes were also taking place. The competition for college spots was increasing as high school graduation became standard and a new kind of credential was sought out by the socially and economically advantaged. Women also began competing for college spots in greater numbers, as did students of colour.

A relatively universal system for grading in the United States reflecting roughly consistent standards, and conveying basic information both internally and externally had a number of benefits—largely in line with the modernist aims of late-nineteenth and early-twentieth century system-builders. It also, however, had a number of unintended consequences.

One consequence was that it stripped teachers of much of their autonomy, with many feeling unable to manually override the grading system. A 1971 survey, for instance, found that only 16% of teachers believed that letter grades were the best reporting method for elementary students and 35% believed that to be the case for secondary education (National Education Association 1971: 81–82). But many felt locked into a widely accepted programme. Many resisted by refusing to forcibly spread their students across a normal distribution, many provided students with multiple opportunities to raise their grades and many issued grades for different kinds of work, not all of it academic. Still, they graded. How could they not, with so many depending on grades for movement through the educational system?

Another consequence was that students learned how to pursue rewarding grades while minimizing effort. The 'entrepreneurial student', Richard Haswell would later write, shops 'for bargain courses, encouraged by a faculty whose jobs are defined by "course load", administrators who deal in credit hours as if they were coin, [and] institutions whose corpus evolves steadily into the corporate' (Haswell 1999: 284). Students often gamed the system by dropping difficult classes, seeking out particular teachers, or taking 'easy A' courses. To put it another way, the importance of grades as a currency for moving through the educational system had partly superseded the pedagogical purpose they continued to serve. If learning sometimes had to occur along the way, so be it; but otherwise, students would do the least amount of work possible in order to attain the token of highest value.

Though limited largely to higher education, a consumer-based approach also played an increasing role—heightening the focus on student satisfaction. As colleges and universities increasingly relied on student-based evaluations of instructors, they created incentives for grade inflation. As Rojstaczer and Healy (2012) would later write, 'in the absence of oversight, and because of the presence of positive incentives to give artificially high grades, higher education has gradually abandoned its grading standards'. The case was slightly different at the high-school level, where the desire to keep students satisfied was not driven by the fear of evaluations, but the desire to please still mattered significantly.

This kind of 'grade inflation' was not just a function of the entrepreneurial student; it was also a function of the increasing embeddedness of the school system in the society at large and the pressure this interconnectedness could create. For example, during the Vietnam War, those enrolled in college could apply for a draft deferment. But in order to keep this deferment status, a student had to remain in good academic standing. Scholars at the time and since have noted that such practices not only incentivized students to enroll in college (Card and Lemieux 2001), but also introduced an explicitly political component into a student's grades (e.g. Suslow 1976). To fail a student during this period was no longer just a message that the student's work in class had not been up to par, but also a statement about his draft eligibility. Indeed, there is evidence that faculty members grasped the suddenly broader implications of their grades and became more lenient in assigning them (Bejar and Blew 1981; Birnbaum 1977).

Even after the war ended, external considerations continued to be a factor in student grades. Whether it was concern for the effects of grades on student self-esteem (e.g. Bachman and O'Malley 1977; Marsh 1990), intra-departmental concerns about attracting students through the implicit promise of good grades (Becker 1997; Freeman 2010), or concerns, particularly strong at elite schools, about the effect of grades on graduates' job prospects (e.g. Dickson 1984), the trend of grade inflation continued. As of 1975, it was reported that one-half to two-thirds of the marks given in American colleges and universities were As and Bs (Davidson 1975: 122–125). A similar trend was found in high schools (Ferguson and Maxey 1975).

By the 1960s and 1970s, a number of figures had begun making a case that grades were not promoting learning. The primary argument, as voiced in the 1962 book *Perceiving, Behaving, Becoming* (Combs), was for a more humanized system. Extrinsic motivation failed to effectively focus students on the process of learning, or so the argument went, and instead, created unhealthy levels of competition and anxiety. Others agreed. As one campus group put it in 1968: grade ‘pressure’ was ‘more fierce than most faculty members realize [and] ... a corollary is the obvious, continued annoyance of grades as a goad to the attainment of success in courses’ (University of Illinois 1968: 1). According to a study from that period, a majority of students agreed ‘that an emphasis on grades encourages cheating, restricts study to material likely to be on the test, and encourages students to conform on tests and in the classroom to the instructor’s views and opinions’ (Leslie and Stallings 1968: 5). As Sidney Simon wrote in the introduction to a 1976 edited volume *Degrading the Grading Myths*, ‘over the past 50-plus years, millions of students have been systematically wounded by the grading and marking system’. As he concluded: ‘it is time to change that system’ (4). And as Norman Chansky found in 1962, many students responded to anxiety around grades by withdrawing, and responded to low marks with a sense of self-defeat (347–352).

Concerns about validity and reliability also continued.³ As Chansky reported in 1964, grades continued to represent different things to different teachers (95–99). And as a Temple University study (1968) found, different professors teaching the same course tended to produce dramatically different grade distributions.

But what could be done? Alternatives like pass/fail grading, mastery learning and contract grading would not allow for the kind of systems and legibility that grading had in many ways been designed to produce. They would make it more challenging to measure students against each other to convey information in a simple and efficient manner, and to treat education as a uniform market. Some certainly resisted grades. Evergreen State University, for instance, or a number of private K-12 schools created narrative evaluation schemes as alternatives to grading. Yet such efforts are boutique responses that at least appear to be untenable as systemic responses, and some have criticized these schools for failing to offer credits that will transfer (Lee 2011).

While one critic made the case that changing grading systems was necessary because of how ‘integrally related to almost every other aspect of a school’s functioning’ they were, the opposite case might just as easily have been made (Kirschenbaum 1976: 111). Schools and schooling had come to depend on a standardized grading system—for motivating students, for determining placement, and for communicating something about student learning both internally within a school and externally to parents and other interested parties.

Conclusion

Before grading, communication about a student’s work was without short-hand; it required full-length communication between teacher and

pupil. This was a reasonable system for individual tutoring or for work with small groups of students. But the ability for that information to travel or be communicated to those not intimately involved with the learning process—teachers, students, parents—was severely limited. As the scale of the education system became larger and more complex, the limitations of these early forms of grading became more acute.

Early proto-grading schemes largely took the form of honorary titles or class rankings, serving as markers to audiences within a school community of each student's relative merit. By the nineteenth century, some improvization was taking place. From the grammar school to the college, teachers and administrators experimented with numerical systems, class rankings, and other forms of communicating students' academic achievements. Still, these systems were primarily suited to the internal organization of schools. They were relatively low-stakes affairs that often served to channel student energies in the direction of learning.

But as reformers worked to develop a national school system in the late nineteenth century, they saw grades as useful tools in an organizational rather than pedagogical enterprise—tools that would facilitate movement, communication, and coordination. These actors placed a premium on readily interpretable and necessarily abstract grading systems. And, the shift to an external communication device traded the meaning and nuance of the local context for the uniformity and fungibility of these more portable forms.

During the inter-war period in the early twentieth century, reformers began to emphasize a new approach to the challenge of building systems. They continued to seek coordination as college enrollments grew, as well as to promote standardization. But with the presumed advances of the mental testing movement, they also sought to improve the degree to which the grades served as objective assessments of student achievement as opposed to the traditional 'subjective' teacher assessments. Increasingly, school administrators sought to develop or reproduce grading systems that squared with ostensibly scientific approaches to measurement, accelerating the isomorphic evolution of once-disparate grading systems.

By the time soldiers returned from the Second World War, they returned to a fully modern educational system, at least where grades were concerned. Middle and high schools used grades to track students. Colleges used high-school grades in admissions decisions. Businesses were interested in the grades of the graduates they were hiring. And as a result of all this, schools not only had to offer grades as a means of extrinsic motivation for students compelled to attend, but also had to ensure that their grades provided a readily interpretable message to future teachers, schools, and employers about the quality of the student. The implication was that grades were an accurate measure of both aptitude and achievement.

Grades by the mid-twentieth century allowed for a great deal of information to be communicated in a highly efficient way. Not surprisingly, however, they sent incomplete messages. Grades were often arrived at arbitrarily or unfairly. They motivated some but turned-off others. And as with so many forms of external validation, they started becoming ends in themselves. Students learned to game the system; 'grade grubbing' and

‘brown nosing’ entered the lexicon. The ‘gentleman’s C’ became a functional concept among the privileged. Grade inflation began to occur, and an anti-grading movement began to emerge.

The anti-grading movement was, in some sense, a logical outcome of the use of grades in system-building. After all, in order for grades to be useful as tools for systemic communication—allowing for national movement, seamless coordination, and seemingly standard communication to parents and outsiders—they had to be simple and easy to digest. Yet that set of characteristics often conflicts with learning because the outcomes of learning are inherently complicated and messy. Consequently, while grades sometimes promote learning, they often promote an entirely separate set of behaviours.

The upshot of all of this is that educators are stuck in a bind. Many continue to see the potential usefulness of grades as internal signals within a school to communicate with students. At the same time, however, those internal signals reverberate well beyond the classroom wall—the product of a relatively unified system in which grades have significant legitimacy as external signals for communicating to those outside the school. As such, educators continue to use grades to communicate with students about their performance in a single class, but are often concerned about the degree to which they are also affecting a student’s future outside that isolated context. Many, consequently, have compressed grades into a narrower and narrower bandwidth. Critics, concerned with systems, call this grade inflation. Yet it might equally be thought of as an attempt to remove the amplification of a grade, which though issued in a single class, nevertheless echoes across a student’s permanent record. Teachers intend to whisper to their students messages like ‘this is not up to par’ or ‘work a little harder’ or ‘see what your classmates are doing?’ But instead, their messages become roars. This is exactly what reformers have long worked for—a system in which messages travel with great volume and clarity. But such a reality presents real challenges for the processes of teaching and learning. And so educators have made the B-minus, the new D.

This kind of temporary truce between the pedagogical function of grading and the systemic function of the practice will hold as long as the history of the F still lingers, and as long as there are holdouts still employing the full range of the grading scale. That will eventually change, however, and when it does, educators will have to find a new way of reconciling the two functions. They must find a way to work within a system that is universally accepted—one essential for national movement, seamless coordination, and seemingly standard communication to parents and outsiders. And, at the same time, they must find a way to keep students focused on learning and not merely on a set of measurable outcomes loosely connected to the process of education.

Notes

1. Here, we follow James Scott’s use of the term ‘legibility’ by which he means the ability for information to have a universal meaning and understood not just in the local context but by distant observers as well.

2. A similar process occurred in non-natural products in other fields as well. On the rise of Moody's Credit Agency, see: Carruthers (2011, May). The Economy of Promises: The Origins of Credit Rating in nineteenth-century America. Presentation to the Department of Sociology, Stanford University.
3. In 1912, Starch and Elliott published 'Reliability of the Grading of High School Work in English' in which they reported that grading was highly inconsistent across subjects and instructors.

References

- Ames, V. (1943) Factors related to high-school achievement. *Journal of Educational Psychology*, 34(4), 229–236.
- Ashbaugh, E. J. and Chapman, H. B. (1925) Report cards in American cities. *Educational Research Bulletin*, 4(14), 289–293.
- Bache, A. D. (1855) Remarks at the opening of the fifth session of the American Association for the Advancement of Education. *Proceedings of the Fifth Session of the American Association for the Advancement of Education* (Hartford, CT: Tiffany and Co.).
- Bachman, J. G. and O'Malley, P. M. (1977) Self-esteem in young men: a longitudinal analysis of the impact of educational and occupational attainment. *Journal of Personality and Social Psychology*, 35, 365–380.
- Bagg, L. H. (1872) *Four Years at Yale* (New Haven, CT: Charles C. Chatfield).
- Barnard, H. (1854) *Practical Illustrations of the Principles of School Architecture* (Hartford, CT: Case, Tiffany and Co.).
- Beck, H. P., Rorrer-Woody, S. and Pierce, L. G. (1991) The relations of learning and grade orientations to academic performance. *Teaching of Psychology*, 18, 35–37.
- Becker, W. E. (1997) Teaching economics to undergraduates. *Journal of Economic Literature*, 35(September), 1347–1973.
- Beggs, V. L. (1936) Reporting pupil progress without report cards. *The Elementary School Journal*, 37(2), 107–114.
- Bejar, I. L. and Blew, E. O. (1981) Grade inflation and validity of scholastic aptitude test. *American Educational Research Journal*, 18, 143–156.
- Birnbaum, R. (1977) Factors related to university grade inflation. *Journal of Higher Education*, 5, 519–538.
- Bixler, H. H. (1936) School marks. *Review of Educational Research*, 6(2), 169–173.
- Boston Monitorial School (1826) *American Journal of Education*, 1(1), 32–42.
- Bolton, F. B. (1947) Value of several intelligence tests for predicting scholastic achievement. *The Journal of Educational Research*, 41(2), 133–138.
- Bristed, C. A. (1874) *Five Years in an English University*. 3rd ed. (New York, NY: G.P. Putnam and Sons).
- Burke, R. J. (1968) Student reactions to course grade. *The Journal of Experimental Education*, 36(4), 11–13.
- Butler, R. (1988) Enhancing and undermining intrinsic motivation: the effects of task-involving and ego-involving evaluation on interest and performance. *British Journal of Educational Psychology*, 58, 1–14.
- Campbell, A. L. (1921) Keeping the score. *The School Review*, 29(7).
- Card, D. and Lemieux, T. (2001) Going to college to avoid the draft: the unintended legacy of the Vietnam War. *The American Economic Review*, 91 (2), 91–102.
- Carruthers, B. (2011, May) The economy of promises: the origins of credit rating in 19th-c. America. Presentation to the Department of Sociology, Stanford University.
- Carruthers, B. G. and Stinchcombe, A. L. (1999) The social structure of liquidity: flexibility, markets, and states. *Theory and Society*, 28(3), 353–382.
- Cattell, J. M. (1910) The achievements and shortcomings of the American college. *The School Review*, 18(6), 369–383.
- Chansky, N. M. (1962). The x-ray of the school mark. *Educational Forum*, March, 347–352.
- Combs, A. (ed.) (1962) *ASCD Yearbook: Perceiving, Behaving, Becoming: A New Focus for Education* (Alexandria, VA: Association for Supervision and Curriculum Development).

- Chansky, N. M. (1964) A note on the grade point average in research. *Educational and Psychological Measurement*, 24(1).
- Compère, M. (1985) *Du Collège au Lycée: 1500–1850* (Paris: Gallimard-Juliard).
- Creasy, E. S. (1850) *Eminent Etonians* (London: Richard Bentley).
- Cronbach, L. J. (1975) Five decades of public controversy over mental testing. *American Psychologist*, 30(1), 1–14.
- Cronon, W. (1992) *Nature's Metropolis: Chicago and the Great West* (New York, NY: W. W. Norton).
- Davidson, J. F. (1975) Academic interest rates and grade inflation. *Educational Record*, 56, 122–125.
- Dickson, V. A. (1984) An economic model of faculty grading practices. *Journal of Economic Education*, 6(Summer), 197–203.
- Dondlinger, P. T. (1908) *The Book of Wheat: An Economic History and Practical Manual of the Wheat Industry* (New York, NY: Orange Judd Company).
- Dustin, C. R. (1926) A scheme of objective marking. *Educational Research Bulletin*, 5(2), 28–31, 40–41.
- Educational News and Editorial Comment. (1912) *The Elementary School Teacher*, 13(2), 57–65.
- Epstein, S. and Timmermans, S. (2010) A world of standards but not a standard world: toward a sociology of standards and standardization. *Annual Review of Sociology*, 36, 69–89.
- Ferguson, R. L. and Maxey, E. J. (1975) Trends in the academic performance of high school and college students. Paper presented at the 50th Annual Convention of the American College Personnel Association, Atlanta, Georgia, March 5–8.
- Finkelstein, I. E. (1913) *The Marking System in Theory and Practice* (Baltimore, MD: Warwick and York).
- Fischel, W. A. (2009) *Making the grade: the economic evolution of American school districts* (Chicago, IL: University of Chicago Press).
- Freeman, D. G. (2010) Grade divergence as a market outcome. *Journal of Economic Studies*, 30(4), 344–351.
- Gerhard, D. (1955) The emergence of the credit system in American education considered as a problem of social and intellectual history. *AAUP Bulletin*, 41, 647–688.
- Geyer, G. H. (1938) Westwood high goes off the 'gold-standard' marking system. *The Clearing House*, 12(9), 530–533.
- Gladfelter, M. E. (1937) Status and trends of college-entrance requirements. *The School Review*, 45(10), 737–749.
- Goldin, C. (2006). School enrollment and pupil-teacher ratios, by grades K-8 and 9-12 and by public-private control: 1869-1996. Table Bc7-18. In S. B. Carter, S. S. Gartner, M. R. Haines, A. L. Olmstead, R. Sutch, and G. Wright (eds), *Historical Statistics of the United States, Earliest Times to the Present: Millennial Edition* (New York, NY: Cambridge University Press). <http://dx.doi.org/10.1017/ISBN-9780511132971.Bc1-509> (accessed April 10, 2013).
- Hadley, S. Trevor (1954) A school mark-fact or fancy. *Educational Administration and Supervision*, 40, 305–312.
- Hansot, E. and Tyack, D. B. (1986) *Managers of Virtue: Public School Leadership in America, 1820–1980* (New York, NY: Basic Books).
- Harvard University. (1831) *Annual Report of the President of Harvard University to the Overseers on the State of the University for the Academic Year 1830–1831* (Cambridge: Harvard University).
- Harvard University. (1869) *Annual Report of the President of Harvard University to the Overseers on the State of the University for the Academic Year 1869–1870* (Cambridge: Harvard University).
- Harvard University. (1876) *Annual Report of the President of Harvard University to the Overseers on the State of the University for the Academic Year 1875–1876* (Cambridge: Harvard University).
- Harvard University. (1878) *Annual Reports of the President and Treasurer of Harvard College, 1877–8* (Cambridge: Harvard University).

- Harvard University. (1889) *Annual Reports of the President and Treasurer of Harvard College, 1888–9* (Cambridge: Harvard University).
- Haswell, R. H. (1999) Grades, time, and the curse of course. *College Composition and Communication*, 51(2), 284–295.
- Hill, L. D. (1990) *Grain Grades and Standards: Historical Issues Shaping the Future* (Urbana, IL: University of Illinois Press).
- Hinsdale, B. A. (1906) *History of the University of Michigan* (Ann Arbor, MI: University of Michigan Press).
- Kaestle, C. (1983) *Pillars of the Republic: Common Schools and American Society, 1780–1860* (New York, NY: Hill and Wang).
- Kirschenbaum, H. (1976) The change agent in grading reform. In J. A. Bellanca and S. B. Simon (eds), *Degrading the Grading Myths: A Primer of Alternatives to Grades and Marks* (Washington, DC: Association for Supervision and Curriculum Development), 111–115.
- Kirschenbaum, H., Napier, R. and Simon, S. B. (1971) *Wad-ja-get?: The Grading Game in American Education* (New York, NY: Hart).
- Kliebard, H. M. (1987) *The Struggle for the American Curriculum, 1893–1958* (New York: Routledge, 1989).
- Kohn, A. (1999) *Punished by Rewards: The Trouble with Gold Stars, Incentive Plans, A's, Praise, and Other Bribes* (Boston, MA: Houghton Mifflin).
- Krug, E. A. (1969) *The Shaping of the American High School, 1880–1920* (Madison, WI: The University of Wisconsin Press).
- Kvaraceus, W. C. (1939) The primary report card. *The Elementary School Journal*, 39(10), 747–750.
- Labaree, D. F. (1988) *The Making of an American High School: The Credentials Market and the Central High School of Philadelphia, 1838–1939* (New Haven, CT: Yale University Press).
- Lee, D. (2011, June 13) Waste your tuition at Evergreen State College! It's fun! Get high! *Seattle Post-Intelligencer*. Available online at: <http://blog.seattlepi.com/theethicist/2011/06/13/waste-your-tuition-at-evergreen-state-college-its-fun-get-high>
- Leslie, E. K. and Stallings, W. M. (1968) *Student Attitudes toward Grades and Grading Practices*, University of Illinois Office of Instructional Resources, Research Report 276.
- Mann, H. (ed) (1845) *The Common School Journal for the Year 1844* (Boston, MA: William B. Fowle and N. Capen).
- Mann, H. (1846) *Ninth Annual Report* (Boston, MA: Dutton and Wentworth, State Printer).
- Marsh, H. W. (1990) The causal ordering of academic self-concept and academic achievement: a multiwave, longitudinal panel analysis. *Journal of Educational Psychology*, 82, 646–656.
- Meyer, M. (1911) Experiences with the grading system of the University of Missouri. *Science, N. S.*, 33, 661–667.
- Milton, O., Pollio, H. R. and Eison, J. A. (1986) *Making Sense of College Grades* (San Francisco, CA: Jossey-Bass).
- National Education Association (1971) Reporting pupil progress to parents. *National Education Association Research Bulletin*, 49(October), 81–82.
- National Education Association (1974) *What Research Say to the Teacher: Evaluation and Reporting of Student Achievement* (Washington, DC: National Education Association).
- Parker, F. W. (1902) An account of the work of the Cook county and Chicago normal school from 1883–1899. *The Elementary School Teacher and Course of Study*, 2(10), 752–780.
- Peabody, A. (1888) *Harvard Reminiscences* (Boston, MA: Ticknor).
- Pierson, G. W. (1983) *A Yale Book of Numbers: Historical Statistics of the College and University 1701–1976* (New Haven, CT: Yale University Press).
- Pulfrey, C., Buchs, C. and Butera, F. (2011) Why grades engender performance-avoidance goals: the mediating role of autonomous motivation. *Journal of Educational Psychology*, 103, 683–700.
- Rojstaczer, S. and Healy, C. (2012) Where A is ordinary: The evolution of American college and university grading, 1940–2009. *Teachers College Record*, 114(7), 1–23.
- Russell, W. (ed). (1826a) Grammar school of Glasgow, Scotland. *American Journal of Education*, 1 (9), 571–572.

- Russell, W. (ed). (1826b) Mr. G.F. Thayer's school, Boston. *American Journal of Education*, 1 (9), 561–566.
- Rutt, N. E. (1943) Grades and distributions. *National Mathematics Magazine*, 18(3), 120–126.
- School Records. (1864) *Massachusetts Teacher*, 17, 58–65.
- Schudson, M. S. (1972) Organizing the meritocracy. *Harvard Educational Review*, 42(1), 34–69.
- Scott, J. C. (1995) *Seeing Like a State* (New Haven, CT: Yale University Press).
- Searby, P. (1997) A history of the University of Cambridge. In C. N. L. Brooke (ed), *A History of the University of Cambridge*, vol. 3 (Cambridge: Cambridge University Press), 1750–1870.
- Segel, D. (1936) To mark or not to mark, an unsolved problem. *School Life*, 22(2).
- Shearer, W. J. (1899) *The Grading of Schools* (New York).
- Shoveller, J. (1824) *Scholastic Education* (Portsmouth: D.R. Price).
- Simon, S. B. (1976) An overview. In S. B. Simon and J. A. Bellanca (eds), *Degrading the Grading Myths: A Primer of Alternatives to Grades and Marks* (Washington, DC: Association for Supervision and Curriculum Development).
- Smallwood, M. L. (1935) *An Historical Study of Examinations and Grading Systems in Early American Universities*. Harvard Studies in Evaluation, vol. 24 (Cambridge, MA: Harvard University Press).
- Sokal, M. M. (ed.) (1987) *Psychological Testing and American Society: 1890–1930* (Piscataway, NJ: Rutgers University Press).
- Snyder, T. D. (ed.) (1993) *120 Years of American Education: A Statistical Portrait* (Washington, DC: National Center for Education Statistics).
- Starch, D. and Elliott, E. (1912) Reliability of the grading of high school work in English. *School Review*, 21, 254–295.
- Steffes, T. (2012) *School, Society, and State: A New Education for a Modern Age* (Chicago, IL: University of Chicago Press).
- Stiles, E. (1901) *The Literary Diary of Ezra Stiles: 1782–1795. Vol. 3* (New York, NY: Charles Scribner's Sons).
- Stowe, C. E. (1838) *Report on the Elementary Public Instruction in Europe* (Boston, MA: Dutton and Wentworth).
- Sumner, R. G. (1935) What price marks?. *Junior-Senior High School Clearing House*, 9(6), 340–344.
- Suslow, S. (1976) *A report on an interinstitutional survey of undergraduate scholastic grading 1960s to 1970s* (Berkeley, CA: Office of Institutional Research, University of California).
- Temple University. (1968) *Report of the College of Education Ad Hoc Committee on Grading Systems* (Philadelphia, PA: Temple University Press).
- Thayer, G. F. (1856) Letters to a young teacher. *American Journal of Education*, 2, 391–398.
- The credit system. (1862) *Massachusetts Teacher*, 15.
- Tyack, D. B. (1977) The man nobody liked: toward a social history of the truant officer, 1840–1940. *American Quarterly*, 29(1), 31–54.
- Tyack, D. B. and Cuban, L. (1995) *Tinkering toward Utopia: A Century of Public School Reform* (Cambridge, MA: Harvard University Press).
- University of Illinois. (1968) Campus Report (June 1).
- Van Herwerden, P. J. (1947) *Gedenkbok van het stedelijk gymnasium te Groningen* (Groningen: Wolters).
- Veblen, T. (1918) *The Higher Learning in America: A Memorandum on the Conduct of Universities by Business Men* (New York, NY: B.W. Huebsch).
- Wechsler, H. S. (1977) *The Qualified Student: A History of Selective College Admission in America, 1870–1970* (New York, NY: Wiley-Interscience).
- Weld, L. D. (1917) A standard of interpretation of numerical grades. *The School Review*, 25(6), 412–421.
- William and Mary College. (1817) *Book of the Proceedings of the Society of William and Mary College* (Williamsburg, VA: William and Mary College).
- Wilson, M. (2009) Responsive writing assessment. *Educational Leadership*, 67(3), 58–62.

- Wofford, K. V. (1938) *Modern Education in the Small Rural School* (New York, NY: Macmillan).
- Woodbridge, W. C. (1835) *American Annals of Education* (Boston, MA: William D. Ticknor).
- Worlton, J. T. (1932) Shall we eliminate the comparative marking system from the report card?. *The Elementary School Journal*, 33(3), 176–184.
- Wright, B. D. (2005) A history of social science measurement. *Educational Measurement: Issues and Practice*, 16(4), 33–45.
- Yale Ad Hoc Committee on Grading. (2013) Preliminary Report. Available online at: <http://yalecollege.yale.edu/sites/default/files/Ad%20Hoc%20YC%20Grading%20Committee%20Report%20-%20Feb%202013.pdf>