

**The Origins of Academic Standards and Promotion Policies in the Public
Schools**

**William J. Reese
University of Wisconsin-Madison**

Final Report, Spencer Foundation

November 26, 2007

Introduction

Americans disagree about many issues, but they generally agree that standards in the public schools are too low in our competitive global economy. Are schools up to the task? Urban schools seem to be failing the most.

Listen to a few voices of despair.

- According to several reports, some teachers in San Francisco, with a nod from their principal, pass every student despite low achievement. Nearly every student is in at least one grade higher than is merited.
- In New York, Chicago, Baltimore, never mind impoverished New Orleans, reading scores, while rising in some schools, are overall anemic. Given the importance of reading in classroom achievement, this is scandalous.
- Forget about raising every pupil to a proficient level: how can a teacher with pupils from diverse economic, social, and linguistic backgrounds address the multiple factors, largely beyond the control of the school, that shape academic achievement?
- How can the nation's poorest pupils—untold millions live in poverty—achieve more when they attend school sporadically and are taught by the least experienced teachers?

- And, finally, many teachers are doing little more than teaching to the test. Teachers prep their students for exams, pupils go through the motions, even if bored to death—and yet test scores show that many children fall behind. Children memorize material for the test that they barely understand and quickly forget.

No one is surprised by this litany of complaints: that academic standards remain stubbornly low and that mediocrity reigns in many classrooms. But every one of these jeremiads arose in a different world, the late nineteenth century. And they were often registered after, not before, educators spent a generation trying to raise standards through competitive testing and related school reforms. The idea of raising standards, of making all children pass written competitive tests, and of linking promotion to test scores originated in America's cities in the pre-Civil War era. Understanding what happened when educators made written tests more central to school assessment and evaluation will not tell us how to lift academic standards for everyone today, the official goal of federal policy. But it provides perspective on the difficulties involved in moving a very decentralized school system along a common path of reform and school improvement.

Over the last generation, raising standards in America's public schools has become a bi-partisan cause célèbre and has dramatically increased the role of the federal government in educational policy making. Competitive tests dramatically define the nature of modern schooling. Test results on academic subjects help

determine the life chances of children: who will be streamed into academic or non-academic classes, to college prep or special education, and who will gain admission to prestigious colleges or settle for less. Federal initiatives such as *No Child Left Behind* (2002) mandate annual standardized testing in grades three through eight and also shape the curriculum, elevating the status of some subjects over others. Such testing also influences pedagogy, as teachers and administrators tailor instruction to try to raise the average score in every classroom and every school. Critics believe that teachers increasingly ‘teach to the test’ and narrow the purposes of schools to standardized scores. Advocates of higher uniform standards, however, insist that regular testing makes schools accountable, so students, parents, and taxpayers can assess their relative merit.

Historians have the luxury of living vicariously in other times and places, unlike contemporary teachers who face the demands of students, parents, and a sometimes angry public. And the research for this project moved in directions I could not have anticipated when I began thinking about it about five years ago. Beginning in the 1990s, many urban districts began restricting or prohibiting a practice known as “social promotion,” the passing of students from grade to grade irrespective of academic performance. As an historian of education, I knew the standard explanation for the rise of social promotion; it was a familiar one, repeated by scholars and journalists alike. Social promotion, it was said, resulted from the pervasive influence of “progressive education.” The phrase “progressive

education” is rarely defined precisely. But it often serves as a short-hand label to describe the ideas of educators influenced by John Dewey (despite his criticisms of child-centered concepts in various writings, especially *Experience and Education* in 1938). Progressive education conjures up such notions as student-centered pedagogy, the denigration of subject matter and teacher authority, anti-intellectualism and disdain for academic rigor, and letting pupils do as they please. According to many critics, the combined force of these interlocking ideas proved catastrophic over the course of the twentieth-century. In the cold-war, hyperbolic language of its day, the famous Republican manifesto, *A Nation at Risk* (1983), even described the collapse of academic standards in public schools as a form of disarmament, which weakened the economy and undermined national security.

Historians know that cause and effect is never easy to identify and prove, and I’ve never been persuaded by the linkage of “progressive education” and school decline. Few people have a coherent, fixed definition of progressive education, and every generation has produced critics certain that its schools are worse than those of yesteryear. Identifying progressivism, or liberalism, or some similar concept to explain educational ills may provide psychological comfort—for every crime, suspects must be rounded up and justice served—but a generation of historians has cast doubt on the popular association of progressivism and school decline.

When I started my research, I doubted that something called progressivism accounted for modern school practices, including the setting of low academic standards or soft promotion policies. In a book entitled *America's Public Schools* (2005), I underscored how many traditional school practices thrived despite the presumed popularity of child-centered norms among some educational leaders in the twentieth century. Numerous scholars (especially Arthur Zilversmit and Larry Cuban) have also demonstrated that the seeming ubiquity of Dewey's ideals (as followers variously understood them) was not a very good guide to how teachers taught or classrooms functioned. Teachers who embrace the rhetoric of child-centered education have often behaved otherwise; teacher- and textbook-centered classrooms remain common. Teachers often teach as they were taught, and the growing emphasis on federally-mandated standardized testing hardly encourages them to change their behavior by substituting child-friendly pedagogy for drill and didactic methods.

The basic question nevertheless remained: why were the great majority of students, despite their often low levels of academic achievement, promoted from grade to grade in the public schools by the mid-twentieth century? That practice was, I knew, never universal. Several researchers showed that before the 1950s some school districts held many children back compared to nearby communities. Often a handful of schools within a particular school district did so, while higher rates of promotion were usually common in the most affluent neighborhoods.

Since the 1920s, however, the overall national trend was clear: a higher percentage of children were promoted. That trend was partially reversed by the 1980s, especially for racial minorities. Beginning in the 1970s, in the wake of declining S.A.T. scores and evidence that high school graduates were sometimes illiterate, numerous states required pupils to pass minimum competency tests to graduate or earn a diploma. Critics of the testing requirements mandated by “No Child Left Behind” fear that the law will also worsen promotion and graduation rates for children of color in particular, and the poor generally.

After a year or so of reading an extensive array of primary sources on urban schools in the early twentieth century, I concluded that I needed a new strategy to solve the historical puzzle before me. To understand why social promotion became so prominent in the twentieth century, I first needed to know more about the very origins and nature of academic standards and promotion policies. That meant exploring the creation of academic culture within America’s urban school systems in the pre-Civil War North. After all, nineteenth-century administrators and educators first faced the difficult question of how to define appropriate academic standards, teaching methods, means of evaluation, and promotion policies for students in urban schools. How to make schools more accountable and efficient was an abiding concern of reform-minded citizens in the nineteenth century. Cities were the site of innovation, since they had the concentrations of pupils, wealth, and expertise to try to address the complex

issues related to school improvement.

As in every era of dramatic social and economic change, critics then abounded, convinced that schools were falling behind the times, in grave need of reform. Over the course of the nineteenth century, the American economy was transformed by industrialization. The application of scientific principles to machines produced cheaper consumer goods. More uniform, precise measurements became more common in industry. Railroads whizzed goods and people along with ever greater efficiency with the adoption of standard-gauge tracks and coordinated communication networks and schedules. Industrial time discipline, where life was governed by the ticking of the clock, not the setting of the sun, transformed the lives of more citizens, especially in the cities. In a world of such dramatic change, urban schools faced an enormous challenge. Children poured into northern cities from the countryside and from distant lands, comprising a heterogeneous mass that schools, too, tried to make more uniform. But schools are not factories, and children cannot always be molded like pig iron. However attractive to some schoolmen, industrial models of organization and production might be inappropriate for the world of education. Could schools turn a heterogeneous mass of pupils into a common product as well as machines turned out identical shirts and bottle caps? Could urban schools, supported by the wealth generated by commerce and industry, become more efficient and cost-effective? Could they produce an improved product, based on commonly

accepted, standard means of measurement?

The more deeply I read in nineteenth-century urban sources, the more convinced I became that I needed to understand how educators coped with issues of academic improvement from the start. Beginning in the antebellum period in the older northern systems, such as Boston, Philadelphia, New York, and Chicago, public school educators were determined to find ways to raise standards, test pupils, and devise uniform ways to promote pupils from grade to grade. Before I could hope to understand educational policies in the twentieth century, I first had to answer some questions from the founding era of these and other urban school systems. On what basis were children promoted in schools before the early twentieth century? Was social promotion a modern invention or possibly an old practice? What difficulties did administrators, teachers, and children face when schools tried to more systematically measure academic standards and determine who merited promotion?

Spending time in the nineteenth century helps illuminate that world, but it also provides useful knowledge to understand perennial educational issues that continue to inform debates about standards and testing policies. Below I'll sketch out some of what I have learned and will expand into a book on the history of academic standards. Although I've studied a few dozen urban school systems in depth, along with many hundreds of other documents and primary sources, I'll limit my examples to keep this report to a manageable length.

Standards and Promotion

Ours is not the first generation to face the knotty questions associated with raising academic standards, testing pupils, or determining who merits promotion. In 1845, Horace Mann, the nation's most prominent educational reformer, helped conceive of a novel experiment. Together with some members of the Boston school committee, he promoted the use of timed, written examinations, the first ever given in America that compared achievement in different schools. Over the next few decades, such exams were commonplace in most urban systems. Indeed, by the 1870s, city systems regularly examined children in end-of-the-year promotion exams. Competitive written tests used for diverse purposes had quickly spread across urban school districts, especially in the North. In fact, making schools more accountable through testing became so prevalent that it generated a national backlash among educators and many lay critics. By the 1890s, virtually every city had dropped the use of competitive annual exams to determine promotion, though written tests of various sorts remained very common in most urban classrooms.

America's first attempt to raise standards had come to an end, the opening chapter in a continuing saga. Why did the movement occur? How did it shape academic quality in urban schools and affect the lives of teachers and pupils? Why did promotion examinations lose their appeal as the century drew to a close?

While European nations in the nineteenth century created central ministries of education to prescribe curricula and control school inspection and assessment, America followed a different historical trajectory. Compared with the nation-states of Europe, such as Prussia within an emergent Germany, America's federal government had a very weak interest in education. The U.S. Bureau of Education, founded in 1867, was a marginal agency. It gathered statistics, disseminated reports, and published studies as budgets and its tiny infrastructure of civil servants allowed. Education was not elevated to a separate cabinet level post until 1979, and its elimination was part of a campaign promise, never kept, by Ronald Reagan in his 1980 presidential campaign. True to our nation's history, suspicion of the federal role in educational policy remains a familiar flash point in our political system. The Constitution did not make education a responsibility of the federal government, and the Bureau of Education did not employ a cadre of school inspectors, as in England. Neither did state governments nor local districts, where the real power over schooling has traditionally rested. Legally established through the constitutions approved by the individual states, public schools from their establishment were locally controlled, financed largely by property taxes, and governed by lay people on school boards. There were many thousands of school districts in the nineteenth century, and there are still over 14,000 today. The task of testing children, or making schools accountable, was historically a local and state responsibility, without much federal

interference until the late twentieth century.

The absence of any powerful central ministry of education meant that schools typically innovated through imitation and local experimentation. Urban schools—where the first movements toward testing and accountability evolved in the mid-nineteenth century—obeyed the laws of political economy, where a division of labor ruled: board members (whether elected or appointed by city council) thus hired experts such as superintendents; they in turn codified a course of study, wrote manuals of instruction, hired subordinates in the central office, and tested and hired teachers, all subject to board approval. Superintendents shared their knowledge with their peers in other cities through their published writings, correspondence, and regular attendance at professional gatherings, usually at the annual meetings of state teachers' associations or the more elevated world of the National Education Association, established in 1857. Superintendents also championed the adoption of uniform textbooks and wrote (or supervised the writing of) grade-to-grade promotion exams and high school entrance tests, both of which were common before the 1890s. The latter tests were often dubbed the “Olympic games.” Local school boards and superintendents therefore set the standards, which were shaped by customary practices and local political realities but never by state or federal authority. By default, school textbooks framed the official knowledge that constituted the curriculum and the source of knowledge tested during the school year in annual

promotion exams.

English Precedents

America's antebellum reformers saw themselves as progressive and informed by new ideas emanating from Europe, especially in England. American reformers spent considerable time informing themselves about innovations abroad. They did so by reading extensively about developments in England, Prussia, and France and by visiting model schools there or spending time with their educational counterparts across the sea. The English example, which was not followed here, was nonetheless crucial, since England was a basic cultural reference point and had notably experimented with testing. Innovative policies in England were widely discussed by American educators trying to better measure school achievement as well as by civil service reformers trying to undermine the ubiquitous spoils system in Washington.

What did American educators learn from England? Many educational innovations—from Sunday schools to early urban charity schools—were English transplants, which were adapted to local contexts. Written, timed, competitive examinations were a major invention of the nineteenth century, and they, too, can be traced to England. American educators took notice when Cambridge University established honors exams in the early 1800s and when, by mid-century, competitive tests helped place Oxbridge men in the civil service. At a time when tests were expected to help improve governance of the empire,

English politicians also debated and then devised a system of school inspection for elementary schools, achieved under the auspices of Robert Lowe and the Education Department. By 1862, Lowe's department pursued a policy known as 'payment by results,' by which test results helped determine the local school grant from the national government.

The famous literary critic, Matthew Arnold, earned his living as a school inspector, and he as well as the novelist Charles Dickens wrote famous criticisms of this policy that were widely reprinted and discussed by schoolmen on both sides of the Atlantic. Thomas Gradgrind, a central character in *Hard Times* (1854), everywhere became the anti-hero of progressive educators, a pedant who promoted rote memorization and didactic methods and, with his assistant Mr. Choakumchild, terrorized the children. He thus joined another fictional favorite one could love to hate, Wackford Squeers, a sadistic schoolman brought to life earlier in *Nicholas Nickleby* (1839). The imaginative world of Dickens was informed by the author's knowledge of classroom realities. He was a serious student of education and social life, having studied, for example, the tests written by the central government bureaucracy. Dickens wrote extensively about education, testing, and the emerging merit system in numerous articles, ever critical of the utilitarian ideals being imposed on the schools. As complaints about the imposition of standards and tests accelerated in America in the 1880s, critics commonly invoked already familiar Dickensian tropes, of cruel teachers and

overburdened students.

A major, fundamental difference between the English and American experience was thus the relationship between national or federal authority and local schools. Local control over the schools was the norm in America, where reform-minded educators and citizens sought to raise standards without any national intervention or oversight. While leading educators followed the controversies surrounding testing in England, American lacked any tradition of strong centralized government in social life, typified by the withdrawal of federal protection for the ex-slaves in the South by the end of the 1870s. The typical federal employee in the 1870s and 1880s was a postal worker, and postal jobs, always much coveted, were usually awarded to friends and political allies by the party in power. Unlike England, America did not have an overseas empire to defend through its military (the Union Army was dramatically downsized after the Civil War) and politicians usually perceived little need for college-trained expertise in government.

Americans throughout the century remained suspicious of centralized political power, and schools were immune from federal intervention and typically never embraced anything like “payment by results.” With the notable exception of the Regents’ exams established in New York, which helped impose more uniform standards on high schools, the typical state agency responsible for education was weak and poorly staffed. State superintendents of public instruction sometimes

used their office as a bully pulpit for educational improvement (or political advance), but turnover in office was often high and the post lacked much authority over local districts. In an educational world of fairly autonomous school districts, imposing change from the top down could prove difficult. Despite the absence of any central authority on the state or national levels to direct the course of educational innovation, common ideas about testing and accountability nevertheless circulated freely among educational leaders. But reforms had to gain approval district by district, winning support from school board members, administrators, teachers, and, ultimately, parents, too.

Many historians have written about the rise of standardized testing, a momentous movement in educational reform in the early twentieth century. Guided by improvements in statistical measurement and the centrality of the new discipline of psychology in educational research, America's schools were greatly influenced by the invention and mass usage of IQ, achievement, and aptitude tests. These tools of assessment and evaluation were from the start controversial and influential, shaping school policies from ability grouping in the elementary grades to academic tracking in the high school. But competitive tests and concerns about raising standards have an older history, one that was decades in the making and which anticipated many of the major policy issues that remain with us still.

That history is absent in most accounts of the rise of school systems in the

nineteenth century. However, local and state-level school reports, periodicals, and professional meetings routinely advanced the latest ideas on school innovations, from age-graded classrooms to the nature of high school entrance exams. Competitive testing became a mainstay of urban schools, and test results routinely helped determine which pupils were promoted or retained, whether in older cities such as Boston, New York, or Philadelphia or newly established districts from Omaha to San Francisco. Standardized tests were yet to be invented, but debates over how to set standards, measure them, and determine school success and failure preoccupied many school reformers and educators after the 1840s. The movement for standards and new methods of promotion, based in part on test results, moved from city to city, as the latest ideas circulated from east to west, inspired by meritocratic experiments from across the sea.

What Horace Mann Wrought

Written tests began in a modest way in America in the 1820s, when a few seaboard towns established the nation's first public high schools. Most high schools required prospective pupils to pass some form of written examination, a practice that continued in cities until late in the century. Boston established the first high school in 1821. But the most dramatic and well publicized use of competitive tests began in the same city in 1845, when a few reformers on the local school committee conspired with Horace Mann to give a common, timed, written test to the boys in the highest grammar school classes. Without telling the

teachers or pupils about their experiment, a few committee members literally raced, like the white rabbit in *Alice in Wonderland*, from school to school with printed questions and watch in hand to administer the test.

Before examiners made their rounds to all of the schools, some indignant principals, infuriated by the unannounced intrusion, leaked a few questions to the other masters. When the examiners reached one school late in the day, they met a master openly drilling their pupils in the answers! Whether this constituted a moral lapse on the part of the master giving the answers, or was reasonable given the circumstances, was debated in neighborhood meetings, in the press, and at upcoming school committee meetings. The male grammar masters had traditionally functioned like feudal lords over their demesne and were then battling to retain their autonomy from school officials. They were locked in a very noisy public debate with Horace Mann over their right to impose, when necessary, corporal punishment. The tests added more fuel to the pedagogical fires already aflame in Boston.

The examination results embarrassed city leaders, who loved to boast about their model school system. The scores were shockingly low, especially when everyone was reminded that these grammar school pupils were presumably the cream of the crop. “Some of the answers are so supremely absurd and ridiculous,” said one observer, “that one might suspect the boys of attempting to jest with the committee, were it not that there are honest attempts to

trace analogies between words which they did know, and words which they did not.” Many pupils could not remember when the Constitution was written or adopted by the states and had trouble explaining whether the waters of Lake Ontario flowed into Niagara Falls or the other way around. Newspapers scooped the story and ranked the schools by average test score. Competitive tests revealed low achievement but also variation from pupil to pupil, school to school. How could one explain these differences?

It was the first but not the last time that the scores of different schools were compared, teachers and pupils embarrassed, and school principals chastised. The great debate over testing had begun. From the 1840s onward, test results became news and a controversial though common way to measure pupil progress, teacher effectiveness, and school efficiency. City school reports and newspapers regularly published statistics comparing local schools, praising the best and shaming the worst. The question, then as now, remained: why did the scores range so dramatically among schools? Like America’s leading corporation, the railroads, urban schools were supposed to run on time; written tests promised to define common standards of performance and help make schools accountable. Railroads (not the federal government) finally imposed standard time zones on America in the 1880s, and competitive tests similarly tried to bring discipline, order, uniformity, and predictability to the nation’s schools. But they also routinely showed what many students did not know. And if everyone did

well, it was not cause for celebration, since critics then complained that the exam was too easy.

Most teaching throughout Western history had been based on texts, lectures, and didactic instruction, and the new reliance upon tests deepened the importance of written authority. The typical one room, ungraded school that most American children attended in the nineteenth century, however, lacked uniform textbooks, many written tests, or consistent means to measure children's intellectual growth except through oral recitations. If teachers wanted to know what a child knew, they asked a question and waited for an answer. And oral methods of pupil assessment remained the norm in America's country schools. In most of rural America throughout the century, children of all ages attended school together, recited what they memorized from diverse textbooks when called upon, and rarely competed in timed, written exams. Asking a question and hearing an answer: this was how adults from time immemorial routinely communicated with each other, not by responses etched on sheets of paper.

Throughout the nineteenth century, oral methods of evaluation thus remained basic to rural schools. The rise of more written tests in urban schools, which had better though hardly uniform age graded classrooms even late in the century, did not mean that oral methods of evaluation disappeared. Traditionally, schools everywhere held annual exhibitions, where parents came to hear pupils respond to questions posed by the school committee, principals and teachers,

and other esteemed dignitaries. Prize pupils recited poetic and literary standards and read original essays on history or current events. The best of them walked away with a new Bible, a new edition of a literary classic, or some similar prize. Critics often bellowed that these exhibitions, however popular, were frequently a sham; teachers often called on the brightest pupils, who were sometimes prepped with the questions and answers, as had occurred in Boston. A former Bostonian similarly remembered how his high school class in the 1830s had been drilled “day after day” before a public exhibition. When one pupil at the big event was asked to decline a particular word, but not the one promised, he “looked toward the master at first completely dumbfounded, then in utter despair faltered out, ‘That’s not my word, sir!’”

Critics also complained that many students who were unprepared simply skipped the ceremonies, and finding ways to discourage full attendance (or other ways to eliminate the worst testers from the pool) was common enough to raise eyebrows among professionals incensed by this legerdemain. “Dodging the examinations” was how Horace Mann described the behavior of the no-shows. Others said that exhibitions and public examinations were undignified, reminding them of the circus, another popular entertainment. Exhibitions were indeed very popular, and they united proud parents, relatives, and citizens who rejoiced at the progress of the pupils and the local schools. According to some critics, however, these events were “truly delusive spectacles” that inflated “the vanity” of the scholars. In 1836, a writer in

the *Southern Literary Messenger* said that “the desire to be puffed in the newspapers, and talked about in public, is substituted for the love of learning for its own sake, and thereby one of the most important objects of education is greatly obstructed.”

School exhibitions, spelling bees, and other oral methods of taking the academic temperature of a local school nevertheless remained common. Traditional practices had considerable public sanction, and these events helped display how well the schools were doing. Since the earliest days of the republic, Boston’s school children paraded through the streets as part of an annual festival, and students there as elsewhere received medals for academic achievement at popular end-of-the-school-year programs. In Washington D.C., children marched in parades and sometimes received ribbons and various tokens from senators and congressmen, sometimes even from the President. After mid-century, when students had to pass written promotion examinations, neighborhood schools in Chicago, New York, and most cities held exhibitions, parades, and pageants to draw attention to the good educational works in their midst. Thousands of people turned out in some cities after the Civil War to see a handful of high school graduates proudly march across the stage to receive their diplomas, give a commencement speech, sing an uplifting Christian hymn, and otherwise make the community proud.

Many of these practices drew upon diverse and time-honored traditions. In the early nineteenth century, religious organizations and philanthropists funded

numerous infant schools and charity schools for the urban poor. These schools served multiple purposes, functioning as child minders as well as teaching self-restraint, basic literacy, and Christian morality to their charges. Newspapers publicized upcoming ceremonies and editorialized on how the urban poor fared, which usually reflected well on their benefactors. As historian Carl F. Kaestle has argued, urban charity schools evolved into public schools; funded originally by Protestant, non-denominational organizations, charity schools slowly received more public funding and served a broader clientele in northern cities by the 1840s. Elected school boards replaced the charity boards, and cities built high schools and pursued other reform initiatives such as graded classrooms, to upgrade the system scholastically to try to appeal to parents and reach more middle class children.

Moreover, the parading and exhibiting of children before the public, which earlier drummed up donations and public support, drew upon considerable historical precedent. Many of these practices were transplanted from Europe and evolved in their New World context. From Europe came a well known tradition of public events common at least since medieval times involving children and adults, including feast days, harvest days, and religious holidays. St. Nicholas Day, honoring the Catholic patron saint of children, and the remembrance of the Slaughter of the Innocents, by King Herod, had been popular for centuries. So were various examples of charivari, special days when the poor, including unruly

adolescents, had the public sanction to mock elites and authority figures. School celebrations, of course, were expected to be more solemn affairs, and usually were, though unruly crowds at some high school graduation ceremonies caused school officials considerable embarrassment throughout the century.

Marching the urban poor before the multitudes was a common sight in recent English experience and well known to Americans, many of whom had cultural affinities with the former mother country. William Blake, the English romantic poet, describes the annual procession of charity school children in London in poems such as “Holy Thursday” in *Songs of Innocence and Experience* (1794). Similar processions were common in cities on the east coast. Both in England and America, Sunday schools also demonstrated their good works through public examinations of the children, who through their demeanor and response to questions on the Bible revealed their moral and intellectual progress.

Such traditions of public display and community celebration hardly disappeared as the public school system was established and expanded over the course of the nineteenth century. However, the advent of the written exam, in which pupils would be explicitly compared with each other, and with numerical values assigned to their efforts, contradicted the old saw that there was nothing new under the sun. Written and oral worlds, like the skills of reading and writing, were different, though in practice they regularly commingled and often found

ways to inhabit the same classroom. Written tests neither eliminated teachers' subjective judgements nor displaced the importance of oral recitations in pupil evaluations even in urban classrooms that adopted the new style examinations. The world of classrooms, like the world outside, was always eclectic and refused to conform to a single standard, drawing upon tradition and clinging to it while embracing the future.

As early as the 1830s and 1840s, professional educators, especially those based in the cities, tried to make important distinctions between a school examination and an exhibition. An examination, they said, was a private event that followed prescribed rules and determined individual merit in academic subjects. In contrast, an exhibition was a public event, which brought together parents and children, the teachers and the community, to entertain, to inspire, to showcase exemplary scholarship. Schools exhibited superior student compositions, the talents of a local prodigy, or the special gifts of young thespians. The exhibition, said the professionals, resembled the county fair. But it was not an examination per se and could not reveal general school achievement. Such distinctions, however, remained blurry to many people for many decades.

Oral recitations remained the bread and butter of instruction and assessment in all schools in the nineteenth century. But a rising class of northern urban educators emphasized the superiority of written over oral assessments. Just as merchants calculated profits and losses, wrote a contributor to the

Common School Journal in 1839, so must teachers provide a proper accounting of what children learned. Impressions were insufficient, and the “proper examination of the schools settles far-reaching questions.” The author was probably the editor of the journal, Horace Mann, who had no doubt that written examinations were the wave of the future. Over the next few years he championed this innovative way to measure achievement, as the Boston grammar masters would soon discover. According to Mann, the use of printed questions, prepared in the central office and used to compare the work of pupils and teachers in different schools, had obvious advantages over oral recitations. Written tests were impartial, compared with an oral exam; each child was asked the same question, had the same amount of time to answer, and was measured “by the same standard.” Instead of asking a particular child one question—a familiar practice in exhibitions—everyone was asked numerous, identical questions. Every child faced both easy and difficult questions alike.

“Of what use to examine a school,” wrote Mann, “if each boy and girl is to be like Punch and Judy in the puppet-show, and to be told by another the things they are to say?” If teachers could not prep star pupils, they would have to encourage all children to study more, thus lifting standards. And, unlike an oral examination, a written test produced a permanent record, which ended debates over how well a child performed. Written examinations showed what children actually knew, not what pupils, parents, or teachers thought they knew. They

became, said Mann, “a transcript, a sort of Daguerrotype likeness, as it were, of the state and condition of the pupil’s minds,” available “for general inspection.” The work of individual pupils and implicitly, and often explicitly, their teachers could be compared, and the best practices discovered, then imitated, lifting overall achievement. Like the newly-invented camera, the student transcript, report cards, and other written records presumably captured reality, documenting academic success and failure. And someone other than a teacher could mark the exams to ensure objective appraisals of a child’s performance. Thoughts of educational revolution danced in many a schoolman’s head.

Oral recitations nevertheless remained common in all schools, both in the city and in the countryside. But reform-minded administrators and teachers, in urban areas in particular, questioned their value in determining achievement and identifying who merited promotion. Like their country cousins, city schools continued to hold public exhibitions, spelling bees, and other community events late in the century to popularize their schools among parents and taxpayers. But cities from Boston to San Francisco from mid-century onward increasingly tested pupils in written exams. As Mann and numerous reformers emphasized, cities had the concentration of pupils and wealth to build more age-graded, larger schools, allowing for more assessment through timed, written, competitive exams. Doing well on a written test increasingly mattered in the cities, then the leaders in school innovation. Where the urban schools led, hoped Mann and a

steady succession of school reformers, others would ultimately follow.

Written Examinations and Urban Schools

Public schools for city children—tax supported and free and in theory accessible to everyone—were established in the pre-Civil War North, where they ultimately enrolled the majority of school-going pupils. America's urban population was rising rapidly, swelled by internal and foreign immigration. As in the countryside, textbooks formed the course of study and in the primary grades emphasized the basics plus some geography, history, and moral instruction. Secular catechisms, textbooks usually included hundreds of questions and answers and problems for students to ponder. Though the results were imperfect and never fully realized, urban leaders endeavored to build integrated, uniform systems with a standard curriculum and instruction guided by textbooks written for each grade level.

Boston, New York City, Philadelphia, and numerous cities implemented policies to ensure that children were tested regularly: weekly, monthly, end of term, and at the end-of-the-school-year promotion exam. Teachers prepared most of the exams, except for the promotion tests, and usually graded all of them. Without machine readable tests, an innovation of the early twentieth century, the burden of implementing and raising standards—from marking exams to reporting the scores—fell hardest on the children but also landed squarely on the shoulders of the teachers. Teachers continued to call on children to recite their lessons, and

administrators regularly observed how well they conducted recitations, maintained orderly classrooms, and kept their rooms clean and presentable. But the trend was toward more statistical measurement, standardization, and record keeping. This was true of railroads, factories, and other industrial enterprises, common reference points for educators who wanted schools to become more effective and efficient.

City schools published annual reports that bulged with statistics in an effort to show they were keeping up with the times. These reports presented statistics on such matters as attendance, cost of instruction per pupil, monies expended for supplies, salaries, or the cost of a new high school. Written largely by the local school superintendent and his assistants, these reports grew in size, comprising hundreds of pages in the big cities by the end of the century. They frequently included the results, school by school, of promotion examinations, which had become common in urban systems by the 1860s and were usually taken at the end of the spring term. School officials reported that the superintendents and his subordinates typically wrote the promotion test questions and the teachers often graded them under their watchful gaze. The superintendent then stored the exams in his office in case parents complained about why their son or daughter was held back in grade.

While city systems grew at different rates, they all went through similar phases of development. Cities initially had ungraded schools, much like rural

areas. Newly-hired superintendents, however, gave a mix of oral and written tests to the youngest pupils, and written tests alone to the older scholars, to ensure better pupil classification. Based on the test results and teachers' recommendations, the children were usually separated into two broad categories, either to the Primary or Grammar-level divisions. Then separate grades appeared within these divisions, and the highest ones ultimately congealed into a high school that served the oldest pupils, who averaged about 13 or 14 years old when they entered ninth grade. Within a generation or so of expansion, most northern urban systems by the 1870s and 1880s had around 11 or 12 grades, with the majority of pupils concentrated in the first four or five primary (or elementary) grades. This approach to classification brought more children of similar ages together in the same classes, encouraging more efforts at uniform assessment and evaluation. The annual written promotion examinations thus helped to bring order to the chaos urban educators perceived in ungraded classrooms, which remained the norm in the countryside, where most children lived.

A.F. Nightingale, the superintendent in Omaha, Nebraska, typically explained what he found when he arrived there in the early 1870s. "Children, on the first day of the term, flocked to school by fifties and hundreds; many from private schools where they were ungraded; many from public schools; and many still who had not been to school for years." Every urban administrator across the

nation, for decades to come, told a similar story about how the schools enrolled a heterogeneous mass of pupils, of all ages, some well prepared and others not, some ready to learn and others allergic to hard work and self discipline, some destined for success and others doomed to failure. Nightingale saw examinations as the only way to bring order to the fledgling system. He personally gave oral exams to the youngest pupils, who had never taken a written one, and he and his staff gave written tests to everyone else. While parents sometimes wanted their children placed in a higher grade than test scores warranted, the superintendent held his ground and, he said, placed them where they belonged. “The time may never come when parents will believe that other people’s children are almost as bright as theirs,” he feared.

Older cities in the northeast had long used entrance exams to determine admission to high school. To gain permission to even take the exam, pupils in most cities usually needed a note from their grammar school principal, attesting to their moral character and fitness to take the test. That was no guarantee of admission. The entrance test examined pupils in their grammar school subjects and required much regurgitation of rote material. Urban high schools spread with the expansion of public school systems by the middle of the century, though they still enrolled only a fraction of all adolescents. But their growing presence meant that older systems had some familiarity with written tests when Mann famously brought them to the attention of more professionals in the 1840s.

Promoting elementary school pupils through competitive written exams dramatically increased the number of pupils tested, and it required local systems to devise and make explicit common procedures to ensure fairness and bureaucratic order. Codifying standard practices on testing was usually the responsibility of the superintendent, subject to approval by the school board. Board members sometimes intervened to countermand some edicts, but power usually flowed downward from the superintendent to each school principal, who had to explain the test procedures to every teacher, who in turn announced to every pupil the date, time, and rules governing the upcoming promotion exam. In San Francisco, for example, the superintendent's office issued circulars by the 1870s outlining the respective responsibilities of teachers, principals, and pupils, a list of do's and don'ts on test etiquette.

Teachers there as elsewhere usually graded the promotion tests, but not of their own class, and pupils were identified by numbers and not by their names, thus ensuring impartiality. Pupils were warned not to talk during an exam, cheat, or open their textbooks. If they were confused by a question, they should skip it and return to it later. Pupils were also told, "Don't attempt to 'cram' for the examination, for it will only confuse you." And, when taking the exam: "Do not hurry; do not worry; do not get excited and nervous, but quietly write all you know about the subject." One can almost see the beads of sweat gathering on the children's foreheads.

In 1888, a history teacher from Syracuse, New York, conjured up an image of the new classroom reality. “Who of us has not been amused by the ludicrous scene which an examination room presents? Youthful faces are screwed into all sorts of hard knots; hair is made to stand on end, presumably for a free passage of ideas; heads are held together as if to prevent them from bursting.” Some of the pupils looked calm, others terrified, everyone resigned to their fate. “Some are eating pencil-tops, and others seem to be writing with their noses.” Not a few look like they are “on the rack.” Students joked about the creation of some magnificent “marking machine.” The student exams would be “cast into a hopper,” pass “through a series of chambers containing balances, tape-measures, yardsticks, pecks, quarts, and gills,” and then “come forth all accurately gauged and marked according to a uniform system.” Whoever invented such a machine, said the teacher, would make a fortune and spare the teachers countless hours burning the “midnight oil.”

Educators did not have standardized tests or sophisticated apparatus at their disposal. Most promotion tests, like other written tests including high school entrance exams, called for short answers to factual information in the Three R’s, English grammar, geography, and history, all graded by hand. Neither children nor teachers could know how many facts, dates, or rules to memorize for an examination. The tests could be difficult one year, easier the next. It was certainly not the case that there were uniformly high standards that everyone was always

held to in the nineteenth century. If the high school was crowded, for example, entry standards were raised; if relatively empty, lowered. Educators openly talked about this, never pretending that some absolute standard existed that everyone could agree on, year in and year out. Everyone also knew that some graders were difficult, others easy, still others erratic. No one could ensure that pupils were evaluated consistently. Administrators usually lacked training in statistics, still an infant science, so few had sophisticated skills or methods of evaluation to account for why test results varied. So they often simply gathered facts and figures, describing and analyzing trends as best they could. Except in New York State, where high school pupils had to pass a Regents' exam to earn a diploma, standards—linked to test scores on questions based on material from textbooks—were determined and assessed locally.

If students had to adjust to the new pedagogical order, so did the teachers, since administrators everywhere said that pupil scores reflected the quality of classroom instruction. Holding teachers responsible for pupil performance no doubt existed to some degree long before the nineteenth century, but it became much more explicit and routine thanks to the birth of competitive written tests. Even though principals, teachers, and superintendents often noted that pupils hailed from a variety of backgrounds and had different levels of ability or interest in academic achievement, holding teachers responsible for poor performance was very common. Urban administrators not only used the tests to classify and

sort children but also to judge their teachers.

The elevation of numerical scores over a teacher's, or parent's, impressions of how well a child was doing rose to new heights. Objective appraisals in industry depended on the perfection of precise tools that consistently measured material reality reliably, apart from human whim or subjective judgement. Tool and die makers had notably helped the cause in industry. Nothing approaching that sort of mechanical science, however, existed in the toolkit of education. But that did not stop some school administrators from making firm and confident claims about how to interpret test scores or explain why some classrooms did well and others poorly. The teachers were often held accountable. Superintendent Nightingale of Omaha characteristically wrote in 1873 that the "examination papers often gives me more reliable information concerning the real condition of a school, and the real character of instruction given, than could be gained by personal visitation for an entire month. The teachers feel this, and often seem to attempt to cover up their defects in teaching, and to have their pupils promoted, when they know it is the destruction of all good discipline, and the very worst disposition that can be made of the pupil."

While the testing provisions of "No Child Left Behind" anger many teachers today, their forbears were not pleased to hear that they bore the responsibility if children did poorly. Moreover, many critics complained that the new emphasis on written tests—which usually required regurgitating rote knowledge—led teachers to

spend considerable time teaching to them. Teachers felt boxed into a corner. Urban administrators routinely criticized their deadly dull pedagogy—drill and more drill—treating their pupils like parrots as they absorbed the textbook knowledge being dispensed, to be recalled for a written test or recited on command. But elementary school teachers often taught classes with forty to sixty pupils. What alternative was there to cramming them for exams? After all, tests rewarded whoever had memorized the most material from the ubiquitous textbooks. If pupils excused themselves from in-class examinations, as some did, the consequences could be worse than dodging a school exhibition or parade. The final weeks of the spring semester usually found teachers drilling pupils day in and day out, since the fate of the children, and their own reputations, were on the line. Visitors to classrooms after mid-century often found every inch of the blackboard covered with potential examination questions. “What will be on the test?” now became a more pressing concern for many children and their teachers.

Most elementary school teachers had minimal if any professional training, and many had never graduated from high school, though increasingly many in the cities had done so by the 1870s. Compounding the problem of preparing students for exams was the reality that knowledge accumulated at a torrid pace in the nineteenth century. What knowledge was of the most worth was contested and debated on both sides of the Atlantic. What pupils and teachers alike were

expected to know had increased dramatically. The sheer size of textbooks swelled over time. New editions of standard volumes had growing quantities of facts about history, geography, and the natural sciences. In geography there were discoveries that yielded huge quantities of facts about mountains and rivers on many continents. Pupils were expected to memorize the capitals of distant lands and routinely asked in history class to recall the names of battles, wars, treaties, and the achievements of successive presidential administrations. Mastering the basic operations of mathematical computation and the rules of grammar, spelling, and composition remained a challenge for every cohort of students. Whether in written tests throughout the year or in promotion exams, pupils had to remember innumerable facts, dates, rules, and definitions and to solve sometimes complicated problems in arithmetic or false syntax.

Cramming and drill remained common, and critics of the sing-song methods of instruction they encouraged often blamed the tests. One observer in Boston asked: were the schools trying to turn out children like identical clothes pins at a factory? There were alternative paths, not taken. Child-centered theorists, such as the Swiss theorist and educator, Johann Pestalozzi, captured the attention of many progressive educators. He called for more attention to learning from experience and teaching through “objects,” not the abstractions of textbooks, claiming that children learned best through more “natural” methods. But dislodging textbooks and rote memorization from the center of classroom

instruction faced often insurmountable hurdles. If test scores mattered most, pupils and teachers would find the easiest road to the promised land, and object teaching did not typically light the way. If the point of an education was mastery of a surfeit of facts and specified domains of knowledge, drill and aids to the memory seemed more likely to illuminate the path to success.

For parents with disposable income and high ambitions for their children, bookshops also sold “quiz books” and “question and answer books” that were keyed to the leading textbooks. Like the advantages provided today by Sylvan Learning Centers, Kaplan prep tests, and Princeton Review, they gave certain children another way to beat those with fewer family resources on examination day. Some quiz books, now forgotten, went through multiple editions and covered every subject for every grade and had tens of thousands of potential test items. Their authors were often school administrators who obviously knew their market well: the vast majority of the questions required short answers, sticking to the facts, as the martinet Thomas Gradgrind had insisted in his opening speech in *Hard Times*: “Now, what I want is Facts. Teach these boys and girls nothing but Facts. Facts alone are wanted in life. Plant nothing else, and root out everything else.”

Such utilitarian assumptions had guided the world of mass instruction both in England, where ‘payment by results’ became established policy for decades, and in America, with its decentralized network of schools. The dominant learning

theory was faculty psychology, which assumed that the mind was comprised of discrete faculties and, like any muscle, strengthened by use and especially by hard work. Mainstream educators and administrators viewed memorizing facts as the key to disciplining the mind, and written tests served multiple disciplinary functions. Urban administrators used them to measure teacher effectiveness, to lift student achievement, and to help standardize the course of study.

Over the course of the century, written exams challenged the dominance of oral methods and became a familiar aspect of going to school, at least in America's cities. Writing on slates, blackboards, and paper everywhere became more common, even in the countryside, when budgets allowed. The more children wrote down, the more they could be compared. Rule of thumb methods of evaluation based solely on impressions or oral recitations remained common enough but lost their prior legitimacy. Even if the norms varied from year to year, common, locally-agreed upon measures of achievement became as essential to a modern school as standard-gauge rails or interchangeable industrial parts. By setting standards to the leading textbooks that had triumphed in a national market, local administrators also ensured the spread of somewhat common standards across cities, which had more in common with each other than they did with the ungraded rural schools. Historians have long documented the high levels of geographical mobility in the nineteenth century, so the adoption of common textbooks to help set standards may have helped some children keep up with

their studies as their families moved across town or to another city.

The ability to compare the achievement of different children fascinated many educators, dependent upon taxpayer support to finance their schools and always under pressure to modernize and economize. By the 1870s and 1880s, most urban superintendents gathered and periodically published statistics on the ages of pupils in every grade and the promotion rates for various neighborhood schools. But by the 1890s, promotion examinations, once nearly universal in northern urban school districts, had disappeared, undermined by a rising chorus of criticism about their allegedly pernicious influence upon students, teachers, and the classroom. Exactly what influence had the exams had on America's schools? Why did they disappear?

The Forgotten History of “Social Promotion”

The official policy in most urban systems in the 1860s, 1870s, and 1880s was that children ordinarily had to pass an annual, centrally administered examination to merit promotion. What had been true for entrance to most public high schools for decades now became established policy for grade-to-grade promotion. The social history of how policies actually worked reveals a more complicated story. It was one thing to make every child in the urban North take a promotion exam. But the practical and political consequences of failing too many, or the wrong, pupils, quickly became apparent in many places. A prevailing myth in the history of education is that social promotion—pushing students along who

do not merit it—is a twentieth century invention, a misbegotten idea of the modern educational establishment, which worried more about children’s feelings than about their achievement. In reality, urban schools in the nineteenth century passed many children who failed the promotion exams, especially in the lower grades. I’m still studying the magnitude of the problem in a variety of cities. But in some years in some systems virtually entire classes were marched forward while in other situations retention rates, for particular schools within a district, were high.

A few major trends emerged by the late nineteenth century, and this allows some informed speculation of what happened in subsequent decades concerning the history of standards and promotion policies. First, being held back in grade was common enough that every pupil likely knew someone who had experienced it. Children moved with their families so frequently and arrived at school with such varying levels of academic preparation, whether from the countryside or abroad, that the task of creating and maintaining fully age-graded classrooms constituted a huge challenge in the cities. Absorbing so many variously prepared pupils, sometimes in large numbers, helps explain why universal age-graded classrooms were often difficult to create in many northern cities until the 1930s. During World War I, waves of rural families, notably in the South, were displaced from the land and moved to the urban north, though the flow of newcomers to the north from southern and central Europe ended with the passage of stringent immigration

laws in 1924 and 1926. By the 1930s, the widespread use of ability groups resulted in more age-graded classrooms; lower-achieving pupils were grouped separately within a class or in separate classes, held to lower standards. Then more of them could be passed along, keeping up with other pupils their same age. But passing unprepared children or those meeting a less-than-desirable standard, to the next grade, was hardly unknown decades before ability groups became so universal.

Second, while being held back in school was hardly strange or very unusual in the nineteenth century, there were well known examples of children who, having failed a promotion exam, were promoted to the next grade anyway. Wholesale promotions sometimes occurred even though teachers and administrators knew that it depressed academic standards throughout the system and often led to more student failure. No one had yet coined a phrase for this policy, later known as “social promotion.” Then, as now, experienced educators knew that a child having trouble reading in first and second grade who did not correct this deficiency had a greater chance of failure later compared with a pupil who, all things being equal, had better reading skills.

Given all the time, effort, and stress given in the nineteenth century to creating exams, preparing children for them, and marking the results, the practice of sometimes passing students along despite their poor achievement seems especially odd. But it happened for a variety of reasons. Most simply, booming

student populations necessitated the practice. Many cities could not build primary schools fast enough to meet demand in the last half of the nineteenth century. According to historian Jon Teafor, taxpayers financed an urban infrastructure that was often more impressive than then found in Europe. But the many demands on the public purse meant that school supply could not always match pupil demand. Everyone knew that, whatever the rules might say, pressure for a seat at school often forced officials to pass undeserving pupils. Depending on the time and place, some districts might hold back anywhere from 10-20 percent of children in the lowest grades because of their poor academic achievement. In some years, however, everyone seemed to pass, an understandable response to demographic pressures. Moreover, taxpayers whose children could not find a seat in the neighborhood school often pressured school boards, principals, and teachers to pass more children along to open some spaces for others.

If pupils sat in a grade for too many years, principals also concluded that seat time usually did not lead to better performance, a theme affirmed by numerous academic researchers today. Moreover, the pressure on space was greatest in the lowest grades, where most children congregated and where learning problems were usually first detected. To hold children back in the mostly overcrowded primary schools of the post-Civil War decades, when children could hardly expect to receive much individual attention, did not seem sensible. Those who repeated a grade often did not progress in their studies anyway; there was

no delicate way to say it, but they were occupying a much needed seat. So they were sometimes simply moved along or, if unlikely to make reasonable progress soon, placed in segregated, ungraded classes.

Educators in the early twentieth century would later write numerous books and articles about “exceptional,” or special children, but so-called ‘backward’ pupils were already being assigned to segregated classrooms by the 1870s and 1880s in some urban districts. Whether to test children who repeatedly did poorly, and include them in regular reporting on the academic achievement of schools became practical concerns for educators. And so it remains. Expertise did not exist at the time to diagnose learning problems (dyslexia, e.g.) or their treatment, and under-performing children were often given arts and crafts—manual training classes—and low-level academic courses to fill up their time at school if they fell far behind their peers.

While teachers and administrators in the post Civil War era often held pupils back who failed to merit promotion when it was feasible, influential, angry, and noisy parents sometimes put enormous pressure on local school officials and the school board to bend the rules. The records of every city are replete with stories about parents who had an elevated view of their own children’s intellectual capacity and classroom performance. “In spite of all that has been said and done,” wrote Henry E. Shepherd, Baltimore’s superintendent, in 1882, “the fact still exists, that in many classes pupils are found who are not equal to the work of the grade,

and whose presence in the class is an injury to themselves, and a continual drag upon the class and the teacher. Pupils are often promoted who are unprepared to commence the work of the next grade. . . . Principals should have the judgement and firmness to refuse to pass such pupils.” That was sometimes easier said than done.

As superintendent Nightingale had learned in Omaha, parental pressure on teachers to pass undeserving pupils, especially their own, was often intense. Principals told teachers not to buckle under, all the while telling them to get everyone ready for the next grade. It was common knowledge that irregular attendance, poverty, poor health, failure to study, inappropriate behavior, and undiagnosed learning problems hindered many children’s progress. Then, as now, who wanted to tell pupils they faced another year in the same grade? Teachers—low paid and without tenure—took the brunt of the abuse when a child failed. Then, as now, who wanted to teach in the lowest achieving schools, filled with poor children? These schools were usually taught by young women with the least experience. In the 1880s, San Francisco hired a female school inspector, who made the rounds of the local elementary schools, and she reported the obvious: teachers preferred working on Nob Hill than on the Barbary Coast, from which teachers constantly sought transfers. All children, she said, deserved qualified teachers and emphasized that the poor deserved the best. The supervisor even suggested higher salaries for teachers laboring in schools near the wharfs, an early call for battle pay.

Thus the policy of promoting children on the basis of test scores never guaranteed that everyone who failed would be held back, and teachers with experience usually tried to avoid those who were difficult to teach and had obvious, though undiagnosed, learning difficulties or emotional problems. To say this does not condone what was happening but suggests the need to hold teachers and schools to standards reasonable for their time and place. At all points in history, accountability should apply to all members of society, not simply to teachers and pupils, and one's reference point should be the world of values and expectations that existed at the time in question, not to those of some never-never land or golden age that never existed.

Where else in society in the so-called Gilded Age was merit and objective judgements for decision making honored, where only the brightest people were rewarded? The social realities that existed outside of schools in the late nineteenth century does not exactly conjure up the image of a meritocracy, which has historically been a better theory than guide to practice. Teachers and pupils lived in a world, like all worlds, of human imperfections, where favors were routinely conferred on those who had inside connections and family ties. Immigrants and native-born alike often found jobs in factories or businesses not only because of what they knew or could do but often because of who they knew or were related to. People often traveled immense distances, whether from Italy or Ireland, or from rural Georgia or Illinois, to northern cities because someone

they knew had a line on a job for them. Nepotism was rife in many corners of the economy and government.

African-Americans, native Americans, and even well educated white women were systematically excluded from certain schools, jobs, and professions. There were of course exceptions to this rule, where individual merit counted as well as chance. But racial and gender distinctions as well as family ties and patronage were familiar means of social maintenance and social order, basic to Europe and America. Since the time of Thomas Jefferson at least, the schools were expected to rise above society and, in his indelicate phrase, “rake the jewels from the rubbish.” This was inspiring: the notion that schools could weaken if not eliminate family advantages and allow individual merit to receive its due.

One should neither mock the ideal nor confuse it with universal practice. Schools often reflected the imperfections of the larger world. City councils routinely awarded contracts to build schools, roads, sewers, and everything else needed in the urban infrastructure; they were often given to friends, relatives, and political intimates, rarely to strangers in an open market of competition, which often did not exist. Merit was malleable, seen in the eyes of the beholder, and took various forms. The members of city school boards typically represented individual wards in the nineteenth century. Anyone seeking a teaching post knew that it didn't hurt to know the principal, a prominent school official, or the local

ward representative. In addition, many teachers, especially in cities, were products of the local school system.

There is no reason to impugn the motives of all elected officials to recognize that kickbacks were sometimes a mainstay in urban politics. More innocently, perhaps, it was often understood by the major political parties that favors were commonly and justly conferred on the right people. City teachers had to pass written exams administered by the superintendent's office to win an appointment, but there were always more applicants than jobs. Choices had to be made. A niece of the board member from Ward Twelve who passed the test, it seemed, might merit the job more than the otherwise perfectly well-qualified daughter of a neighbor who belonged to the wrong political party. Schools were always, and remain, intensely political, and they should not be held to some utopian standard no one else at the time was expected to follow.

If this seems to apologize for the failings of the schools, it's again useful to examine how the larger society operated, at the highest governmental levels. Consider the use of civil service exams for federal jobs in Washington. Here was a reasonably analogous situation in which, as in the schools, tests promised objective appraisals, so that individual merit, not family or political connections, allowed the best to be identified and rewarded. The story of the civil service system is however an instructive example of how widely theory could diverge from practice. It puts the inconsistent use of test scores for promotion in

America's schools in context.

Politicians were much slower to praise the value of competitive tests than educators. In fact, congressmen of both parties only grudgingly endorsed civil service tests for federal employment. The most reliable support for civil service reform in Washington came from whichever party (often the Democrats) was out of power, to weaken the dominant group's hold on patronage, as well as elite New Englanders (often disaffected Republicans), offended by their own declining political and cultural influence. But civil service tests, which covered an increasing number of workers by 1900, usually only covered high-turnover, lowly clerical positions; the exam standards were deliberately kept low to prevent college graduates from "meriting" all of the jobs. Moreover, politicians who pressed merit systems too far, including the person for whom the Pendleton Civil Service Act (1883) was named, became outcasts among former friends and allies, their political careers often destroyed by members of their own party!

Those entering politics—whether elected to the local school board or to the House of Representatives—assumed that doling out jobs was part of the job. As it had long been said, "To the Victors Belong the Spoils." One especially rapacious representative of the people in Washington, no friend of reform, was even nicknamed "Me Too." Financial scandals during the Grant Administration plus the infamous behavior of Tammany Hall in New York City only reinforced the popular view that politics and corruption were bed mates, a bipartisan love affair. To

ensure that the spoils system was not seriously weakened by the Pendleton Act or by later emendations, congressmen also protected their patronage by ensuring that jobs did not automatically go to applicants with the highest test score.

Republicans, for example, established quotas to guarantee that the spoils would be evenly distributed geographically; a person with a lower test score from a state whose quota of jobs had not been met bumped a person with a perfect score from a state that had already done so—an early example of what was later called affirmative action. The merit rules were routinely undermined by administrative loopholes and political chicanery. Few people in Washington thought test scores should determine everything. There was slippage in the highest ranks of office, as in the neighborhood school. Situations and conditions demanded flexibility and a broad view of merit.

That said, schools often did try to base student promotion largely on competitive exam scores, honoring meritocratic ideals that were often regularly mocked in practice throughout the political system. Girls, for example, often outperformed boys at school, earning higher grades on average and enjoying higher rates of graduation from high school. (Then as now boys are more often retained in grade.) It's difficult to identify many places outside of school in the nineteenth century where the conditions for equal competition or public reward for merit also existed. Moreover, the inconsistent application of promotion exam results in determining grade-to-grade promotion does not diminish the long-range

importance of written examinations of all sorts that reshaped urban schools over the course of the nineteenth century. Children in many cities at different points in time were held back in grade, if their promotion test scores were too low, if conditions allowed, and the political will existed to support the decision. Sometimes school boards backed principals and teachers who were willing to take the heat from disaffected parents.

Competitive school exams, whether to determine monthly or semester grades, or for promotion, likely had many beneficial effects. More than a few observers believed that, despite the obvious stress upon children and students, such exams raised standards, helped calibrate in a rough sense desired levels of achievement for each grade, and ensured that teachers could not simply rely upon their own impressions to determine student progress. If they functioned as a servant and not a master, exams, said their advocates, had a salutary influence.

Interpreting the Score

Competitive tests were always controversial, since the exercise, in the language of the day, often separated the brightest from the dullest. Invidious distinctions made about children naturally infuriated some parents. In addition, the pressure to excel, and surpass others, if taken to extremes, could become competition for its own sake and a terror to some teachers and pupils. And teachers and superintendents did not always see eye-to-eye on the uses of tests. Who was being judged? Like industrial bosses and factory managers, whose job

was to tone down labor militancy and keep the workers in line, superintendents in the big cities were sometimes in a running war with teachers, locked in an obvious power struggle. They could simultaneously use exams to help measure and maintain pupil achievement and blame teachers for poor pupil performance. Teachers faced annual reappointment, so the stakes were high.

Debates over the value of the exams erupted everywhere, but written (as well as oral) tests continued to be given in most cities through the 1880s and used, whenever conditions and politics allowed, to help determine promotion. Complaints about this new regime of testing were by then fairly deafening, much as critics today have sounded off against the testing provisions of *No Child Left Behind*. The teachers, they said, are teaching to the test, the value of a school is reduced to test results, the methods of instruction are distorted, the pupils are bored as they are drilled and drilled some more, even though the test scores don't rise appreciably. How, asked some observers, can we hold our best teachers in the worst achieving schools? Has the intrinsic love of learning that should be instilled in children given way to test scores?

Such complaints filled the pages of teacher magazines and more high-brow venues in the late nineteenth century. As enrollments expanded, there were more written tests to grade, multiplying the chances of producing unhappy pupils and parents. And publicity about low test scores, as Boston's educators learned in 1845, brought unwanted attention to the schools. After the Civil War, the subject

of test scores made national news. Once again, the controversy arose in Massachusetts. In 1878, George A. Walton, an agent of the Massachusetts State Board of Education, conceived a novel plan: a common written exam in the Three R's in the various school districts of Norfolk County, south of Boston. This was the first effort to compare test results across school districts. Walton and his colleagues tested 2,866 primary school pupils (those selected had attended school about four years), and 2,095 grammar school pupils (who had attended twice as long). Test scores varied from school to school, but the pupils recorded an average of 57 percent. Reports circulated that many pupils did not have much experience holding a pencil and that many had never taken a written test. Walton's lengthy report, which included facsimile reproductions of children's handwriting and written responses, was widely circulated among state departments of public instruction and the editors of teachers' magazines. It proved to be unwelcome publicity. The schools of Norfolk County were about to receive a public thrashing.

The Norfolk County report encouraged heated exchanges about whether the nation's public schools were succeeding or failing that soon appeared in the *New York Times*, the *North American Review*, and numerous educational journals. Tales of educational decline abounded. Writing in the *Times*, Richard Grant White, a well-known literary critic, attacked the family backgrounds of the "uneducated stock" with their "slovenly habits of mind and of speech" who filled

city schools. White simply said, “judge the system by its fruits,” and then ask why schools cost so much. It was obvious to him that the public schools turned out lazy pupils: the girls destined to become unreliable domestics and addicted to fashion, the boys inferior in thought and “manners” to an older generation. Continuing his diatribe in the *North American Review*, White argued that crime increased as more money was spent on schools. “Our large towns swarm with idle, vicious lads and young men who have no visible means of support. Our rural districts are infested with tramps—a creature unknown to our fathers, and even with us in our youth.” White urged the elimination of public support for high schools and endorsed more “public farms and public workshops” to straighten out the unruly and unlettered children of the poor. For White, Norfolk county provided telling proof that public schools had failed.

Test scores thus provided an opportunity for public school friends and foes alike to air their views on various educational policies and practices. Some critics blamed the low scores in Norfolk County on the teachers, others blamed the testers. Except for White, few said much about the students. Not surprisingly, Walton concluded that schools should hire more administrators to better monitor the teachers. His employer, the State Board of Education, tried to deflect rising criticisms by announcing that the schools of Norfolk County were doing fine. If the tests were given to schools elsewhere, it claimed, the results would be similar. “The examination exhibits many and great defects in the system of instruction practiced in some of the schools of the county; but it shows an average of

excellence, that, if carefully analyzed, ought to encourage the zealous and nervous friend of our Massachusetts system of public schools.” Few citizens, however, swallowed the notion that an average test score below 60 percent was something to boast about. Who should get credit and blame when the scores were posted? The board had an answer: “The schools deserve credit for their best products; and, as a general thing, they should be held somewhat responsible for their poorest.” Officials thus promoted a double standard: full credit for educators for high scores, only partial blame for low scores. Let the experts interpret the scores and explain their meaning!

Numerous observers concluded that the schools were too soft and too child-centered and needed firmer discipline and teacher authority. Others who favored more child-friendly pedagogy said the tests covered trivia and obscure facts; didactic forms of instruction, they said, had obviously failed, and many children rejected attempts to force them to memorize the surfeit of facts, dates, and rote material common on most exams. Still other critics, forgetting or not knowing that Boston’s choicest youth had done miserably when tested decades earlier, simply saw the low scores as another sign of cultural decline, a fall from those better times when they were in school. The Norfolk County exams, comparing achievement across school districts, produced such negative publicity that they were not imitated anywhere anytime soon.

One fan of child-centered instruction, Charles Francis Adams Jr., wrote

engagingly about the Norfolk County controversy in *Harper's New Monthly Magazine* in 1880. Adams had become famous in the annals of education a few years earlier, when he applauded progressive teaching practices then being introduced into the public schools in Quincy, Massachusetts, under the leadership of Col. Francis Parker. Parker was a charismatic champion of reform, one of the nation's leading advocates of progressive teaching practices. Adams had been a member of the school committee that had lured Parker to Quincy. Adams feared that city schools especially were becoming "huge mechanical educational machines" that resembled a combination cotton mill, railroad, and "model State-prison." Filling pupils' minds with knowledge had reached a ridiculous level. Textbooks continued to grow in size, making the mastery of all the relevant facts about any subject impossible. But the schools persisted in force feeding scholars more facts through cramming, even when test scores showed that there were limits to what the mind could, never mind should, store.

Little heads were crammed with dates, rules, and facts infinitum, which pupils regurgitated as best they could on examination day and promptly forgot. Urban schools had time tables just like the railroads, with the school day divided up into so many minutes for each subject. But to what end? "Mechanical methods could not be carried further," Adams noted. "The organization is perfect. The machine works almost with the precision of clock-work." But the assumption that all children of the same approximate age should consume the same "mental nutriment in equal quantities and at fixed times" to

progress together from grade to grade worked better in theory than in practice. Retention rates, he thought, were often high in urban systems, especially in the lowest grades, and most pupils left school to work by age 12. Adams asked the question, still a favorite among progressives: did schools exist for the sake of the system or for the sake of the children?

Complaints about excessive testing accelerated after the Civil War for the simple reason that written competitive tests had become common and influential. Critics of competitive written exams commonly invoked images of Thomas Gradgrind or Wackford Squeers. A host of physicians as well as lay people testified that excessive study and examination impaired the health of teachers and pupils, especially that of girls and women, a serious concern in the 1870s and 1880s. Poor health could weaken future mothers. Tapping into this many-pronged assault on testing, a teacher from Brooklyn, Carolyn Le Row, wrote a minor best seller in 1887 entitled *English As She Is Taught*, which reprinted ludicrous student answers on school exams. Her book became the century's most famous collection of student bloopers and earned a flattering review from Mark Twain, which helped sales. Like Charles Francis Adams, Le Row championed child-centered teaching methods. She hoped her book and subsequent writings would persuade citizens to eliminate the sing-song methods administrators had forced upon teachers.

English As She Is Taught captured national media attention, and Le Row

continued to write on behalf of progressive education for the next two decades. Since the mid-nineteenth century, newspapers and magazines had occasionally reprinted the silliest test responses by pupils, starting a now venerable tradition of making fun of what students don't know. Jokes often work best when they parody widely accessible, shared conventions. By the time Le Row's book appeared, educated people were very familiar with written examinations and the unintentionally comical answers of some pupils. Readers likely chuckled upon learning that the Eucharist is "one who plays euchre," that "the Britains were the Saxons who entered England in 1492 under Julius Caesar," and that "Abraham Lincoln was born in Wales in 1500." But some prominent citizens and intellectuals writing in national magazines and prominent newspapers concluded that the test scores were low because teaching methods were too lax and pupils were so lazy. This was not what the Le Rows of the world wanted to hear.

Without question, however, the reaction against the exaggerated uses of written tests accelerated in the 1880s. As student populations grew, tests consumed more time from administrators and teachers and money from the school budget. Weary administrators highlighted the burdens of office, calculating the weeks of labor involved in grading, annually and by hand, tens of thousands of promotion exams. Administrators without enough assistants saddled teachers with the job—without any extra pay. The time and expense involved in writing, printing, administering, and grading the tests—and dealing with the fallout when

children failed—grew in tandem with expanding enrollments.

Building upon the pent up anger of teachers in particular, Emerson E. White, a former superintendent in Cincinnati and a nationally prominent educator, assailed promotion tests in 1891 in a widely circulated report published by the U.S. Bureau of Education. White said the tests wasted considerable time and caused pupils and teachers too much anxiety. While he did not oppose written exams, he thought promotion should be based on some combination of daily marks on recitations and monthly, term, and final written examinations. By the time the report appeared, many cities were actually doing just that. Promotion exams were either abandoned or no longer used as the sole measure to determine promotion. By the early 1890s, mounting criticisms from teachers and other citizens in England similarly ended the contentious policy of “payment by results.” Many educators on both sides of the Atlantic promised to slay the ghost of Gradgrind.

Conclusion

The first wave of competitive testing in the nation’s schools, largely centered in the cities, was coming to a close. But written competitive exams, taken in a fixed amount of time and bound by common rules, hardly disappeared, as superintendents and other educators often recognized their value and defended their use. By the end of the century, urban pupils regularly faced many more written tests than had been true before the Civil War, and superintendents

occasionally gave special exams to measure teacher proficiency and pupil achievement. There was no retreating to the methods still being used in ungraded rural schools. Tests gave the superintendent and his staff more power over teachers, and written exams, supplemented with other indicators of classroom achievement, obviously helped teachers and everyone else determine how well children were doing in school. While agreeing that numerical scores on tests were not the only measure of school quality and teacher effectiveness, administrators emphasized in the 1890s that, even if promotion tests declined in influence, written exams were here to stay.

As more and more pupils were compared in written tests, educators armed with the statistical data could prove what everyone had always known: that children were different, that they varied by nature and temperament and family circumstance, and that not everyone was destined to succeed at school. The scores helped reinforce the assignment of pupils to ability groups within classrooms or to separate ungraded classes when pupils performed so poorly that they seemed ill suited to a regular class. What was later known as special education was already becoming visible on a small scale in most urban areas by the end of the century. Numerical scores lent an aura of scientific authority to standards, and they usually highlighted the range of achievement found in any given group of pupils in any given subject. And the test results, then as now, led to much hand-wringing, posturing, and political controversy.

The widespread use of promotion exams and written tests more generally during the school year had helped transform urban schools, the models for the rest of the system as districts consolidated. Uniform written tests helped promote a more unified curriculum. Facing tests regularly undoubtedly forced children to study more, even if critics thought children were too often overwhelmed with the information and often failed to understand what they memorized and regurgitated on tests. The explosion of knowledge—the sheer spread of information on every academic subject—had made teaching ever more complicated and demanding. Into the schools more children arrived, faced with textbooks that brimmed with facts and knowledge whose boundaries ever expanded.

Teachers and administrators struggled to accommodate themselves to the new information age and global economy; then, as now, it was difficult to define what was credible, useful, and pertinent knowledge. Written, timed, competitive tests greeted urban children as they entered school, helping determine who succeeded and failed, and made the grade. We are still testing the children, still arguing over what knowledge matters most, still lamenting the tiresome ways of traditional instruction. And we still wonder what to do with those left behind.