

Testing and Accountability, Historically Considered

Diane Ravitch

Nowadays, one thinks of testing and accountability as twins in education; tests, it is assumed, produce the data on which accountability for results is based. However, as one surveys the history of American education, it quickly becomes clear that the history of testing is far more venerable than the history of accountability. Furthermore, history reveals a fundamental conflict between the education profession and laymen in the way they perceive the uses of testing and accountability. Much of the current controversy over testing and accountability can be traced to this conflict of perceptions.

Testing has long been a staple in American public education. Schools and colleges administered tests of various kinds in the nineteenth century and used them to limit promotion to the next grade and for college admission. But the contemporary idea of accountability, that is, holding not only students but also teachers, principals, schools, and even school districts accountable for student performance, is a recent invention. The idea of measuring the quality of education by the academic performance of students is not one with a long pedigree.

Nineteenth-century schools frequently tested their students to see if they had mastered what they were taught. Some tests were districtwide, whereas others were written by the teachers. The tests were very specific in terms of what they expected students to know; there was little room for ambiguity or nuance. Students who didn't pass the tests in history, geography, and arithmetic were "left back." They were held "accountable" if they failed to learn. In many school districts, students had to pass a test and/or complete a designated course of study in order to enter high school.

It was generally accepted by school officials and parents that enrollment in high school should be for those who could handle the work and that many youngsters either could not or did not want to do so. At the end of the nineteenth century, less than one of every ten adolescents went to high school. Those who wanted to enter high school were expected to demonstrate that they were ready. Most children left school at the end of eighth grade; either their families could not afford to keep them in school or they saw no reason to remain. Attendance was voluntary, and a high school diploma was not necessary for most kinds of work.

The small number of students who wanted to go to college had to prepare for college-level work. Many colleges accepted anyone who applied, but the most prestigious colleges required students to pass their admission examination. Elite colleges, such as Harvard, Princeton, and Yale, informed prospective students about the content of their examination, and students prepared to be examined by the college of their choice. Public and private secondary schools alike prepared their students for these college entry examinations. Many principals and headmasters complained about preparing students for different colleges (each with its own selections of reading from Latin authors). The volume of complaints inspired the creation in 1900 of the College Entrance Examination Board (CEEB), which prepared a single test of admission for a large number of colleges and allowed students to select the courses in which they would be examined.

Thus, those who wanted to go to a good public or private college knew that they would be held accountable for what they had learned. The establishment of the “College Boards” reinforced this expectation because the CEEB published its syllabus in different subjects, teachers taught it, and students were examined on whether they had mastered it.

School teachers in the nineteenth century were often required to pass a test of their knowledge, and often they were interviewed by members of the local school board. Oftentimes, the interview included a close examination of their religious views (the local school committee usually included members of the clergy) to make sure that the prospective teachers harbored no unconventional views. Once teachers were accepted for service, however, there were no more tests of their suitability or capacity. If their students failed to learn, it was the students’ fault, and the students suffered the consequences of their ignorance or their lack of willpower by failing or by dropping out without graduating from high school.

The design and administration of testing began to change in the early years of the twentieth century after the field of educational psychology was established. As a new discipline, educational psychology found an institutional home in the new colleges of education and turned almost at once to the reform of educational testing. Most of the tests written by school districts and teachers were simple tests of recall; the psychologists criticized them for lacking reliability and validity. Their efforts to introduce scientific rigor into testing required standardization and adherence to psychometric principles of scientific objectivity.

The leading educational psychologist in the first half of the twentieth century was Edward L. Thorndike of Teachers College, Columbia University. Thorndike was determined to demonstrate that education could become an exact science; in his own research and in his textbooks, he stressed the importance of applying rigorous scientific methods to school practices. Thorndike eventually became known as

the founding father of modern educational testing; he developed standard scales for testing pedagogical methods and school subjects.

Although Thorndike worked on perfecting tests as measures of academic performance, he had no interest in using testing for purposes of accountability. According to his biographer, Geraldine M. Joncich, Thorndike expected “that the high prestige of science would minimize outside interference, that the indisputability of scientific ‘laws’ would reserve educational control to educators trained in the principles and methods of a scientific pedagogy.”¹ Like other progressives, Thorndike believed that education was a function of the state and that its administration should be a professional matter in which public oversight was strictly limited. By improving professional practice, Thorndike thought that there would be even less reason for noneducators to become involved in the operation of the public schools.

Thus, Thorndike’s work on testing was intended to strengthen the profession, not to increase public oversight of the schools. He wanted to see teaching evolve from an art to a science; toward that end, he promulgated various “laws of learning.” The Law of Exercise implied that children learned better when they practiced what they were supposed to learn. Thorndike’s Law of Effect implied that children would learn better if the act of learning brought them a sense of pleasure or satisfaction. Progressive educators, who deeply admired Thorndike’s intention to put education on a scientific footing, ignored the Law of Exercise, which smacked of rote recitation (which they despised), and applauded the Law of Effect, which supported child-centered schooling. In the child-centered school, the interest of the child was supposed to be the primary stimulus to learning,

¹Geraldine M. Joncich, “Science: Touchstone for a New Age in Education,” in *Psychology and the Science of Education: Selected Writings of Edward L. Thorndike*, New York: Teachers College Press, 1962, p. 7.

instead of the child's effort, and the Law of Effect suggested that children would not learn if they were required to engage in studies that they did not enjoy.

Led by psychologists like Thorndike, the testing movement evolved as the progressive education movement gained ideological dominance of the education profession in the 1930s and 1940s. Professional educators embraced testing because it seemed to place education on a scientific plane where decisions could be made on a professional basis and could withstand the entreaties of parents. Progressive educators were enamored of both child-centered practices and social efficiency; they sought to make the schools less academic and to create multiple programs for children who were not interested in traditional schooling. Tests, most especially intelligence tests, were used to sort children into different curriculum tracks so that the nonacademic students would be correctly placed into vocational and industrial programs. During the years of progressive hegemony, tests were extensively used for determining the aptitudes and intelligence of children and guiding them into the right curriculum track.

During the 1930s, as progressivism was in its heyday, the schools were encouraged by progressive leaders to promote children each year regardless of their performance. This practice came to be known as social promotion. At one level, it was a response to the Depression; it was intended to keep young people in school and out of the job market, thus reserving jobs for adults. But at another level, social promotion was championed by progressive educators who were concerned about the effects of retention and failure on the psychological well-being of the child. Advocates of social promotion insisted that schools should put less emphasis on subject matter, discipline, and grades and more emphasis on children's social adjustment.

Thus, although testing was regularly used in the schools, there was no belief within the profession that tests should be used to hold anyone accountable. The spread of social promotion meant that even students would not be held

accountable for their performance in school. These trends were not reviewed by the public; no polls revealed whether parents wanted their children to be socially promoted. The transformation of the school from a meritocratic institution into a custodial institution happened almost entirely without public participation. These changes were facilitated by the profession's belief that the practice of education was strictly a professional matter that need not involve members of the public other than as taxpayers and by the development of a technical pedagogical lingo that made education seem to be near-incomprehensible to untutored laymen.

Interest in accountability may be traced to the landmark report *Equality of Educational Opportunity* of 1966, known as the Coleman report for its lead author, sociologist James Coleman. Written as a study to compare the distribution of resources and opportunities among children of different races, the Coleman report also examined differences in achievement scores, or outcomes. The study was significant for many reasons, one of which was its shift in research focus from inputs to results, which resulted from the authors' decision to examine how school resources affected achievement.²

Before the Coleman report, education reform had focused solely on the issue of resources, on the assumption that more generous provisions for teachers' salaries, facilities, textbooks, and supplies would fix whatever ailed the nation's schools. After the Coleman report, reformers advanced a broader array of proposals, many of which sought changes in performance rather than (or in addition to) increases in resources. In the late 1960s, some urban school systems experimented with decentralization or community control in an effort to shake up the bureaucratic structures that managed the schools. In 1971, the federal government sponsored a choice program in Alum Rock, California, which allowed

²Diane Ravitch, *The Troubled Crusade: American Education, 1945–1980*, New York: Basic Books, 1983, p. 168.

mainly low-income parents to choose among their district's public schools. During the same period, scores of school districts tried performance contracting with private firms to deliver remedial services to low-income students.

These tentative ventures into decentralization, choice programs, and privatization schemes resulted not from the demands of professional educators but because of the intervention of policymakers and local school boards. Professional educators continued to believe that the inadequacies of the schools could be resolved by adding additional resources. However, policymakers, public officials, and, in some cases, community activists and parents concluded that the problems were structural consequences of the bureaucratic system of public education and had to be addressed by competition or structural change.

The shift in focus from inputs (resources) to outputs (student achievement) was facilitated by the increased availability of test scores. The establishment of the National Assessment of Educational Progress (NAEP) in 1970 provided cumulative new data and trend lines to document the educational achievement of American students; after 1992, NAEP reporting was expanded to include samples of students in participating states. Another source of information about student achievement was contained in international tests of mathematics and science (the two subjects on which students in different nations may be compared); American students in grades eight and twelve, when international tests were usually given, often performed poorly and seldomly above the international mean.

As more and more information accumulated about student performance, elected officials came under pressure to “do something” about low achievement and about the large gaps among different groups of students: between those who were poor and middle-class and among students of different races. Confronted with the need to improve their schools in order to attract new industries to their states and localities, elected officials—especially governors—looked at education much as

they looked at other functions of government and at private corporations. They concluded that what mattered most was results—that is, whether students were learning. They used test scores as the best measure of student learning, and they urged that schools should focus relentlessly on improving student achievement.

By the early 1980s, governors were turning to business leaders as their natural allies in trying to improve their state's educational system. In every state, education was the single biggest budget item, usually consuming 40 percent of the state's expenditures. Some governors wanted to get education under their control, some wanted to make education spending more cost-effective, and most wanted to accomplish both. The governors looked to business leaders for advice on managing complex, labor-intensive organizations. The business leaders looked at the schools through the lenses that were customary for them. They expected to see transparency of reporting about budget, resources, operations, and results; they expected to see accountability for performance. And they encouraged governors and other elected officials to consider incentive structures that worked routinely in business to improve performance.

Thus, over the past generation, a split has occurred between professional educators and the public officials who control the purse strings. In effect, there are two competing paradigms of education reform at work simultaneously and not always harmoniously. Professional educators and their allies in higher education continue to focus on inputs (resources for reducing class size, increasing teachers' salaries, and expanding teacher training, for example), whereas policymakers representing the public seek accountability for results.

If we accept the notion that there are two competing paradigms, then we can see how these paradigms are in constant tension, with advocates of first one, then the other, gaining brief advantage. Policymakers have sought accountability for students, teachers, schools, and school districts; profes-

sional educators have largely resisted these pressures. The grounds for their resistance have varied, depending on the issue, but in every instance the educators have sought to water down accountability and maintain professional discretion.

The policymakers want tests to have stakes for test-takers attached to them so that students will exert greater effort to pass them; the professional educators (with some notable exceptions) seek to soften and eliminate any stakes for students. The most notable exception to this generalization was Albert Shanker, who was president of the American Federation of Teachers until his death in 1996. Shanker advocated standards, testing, and stakes, and his union has mainly continued to follow his line; strong objections to stakes for students have been raised by the larger teachers' union, the National Education Association, as well as by other organizations of education administrators and researchers.

The policymakers have endorsed the standards-and-testing approach, in which states describe what students are expected to learn then test to see whether they have learned what they were presumably taught. Professional educators have gone along with this strategy with varying degrees of enthusiasm but with a chorus that warns about the danger that real incentives and sanctions will cause "teaching to the test," "narrowing the curriculum" to what is tested, and cheating by teachers.

The policymakers want to use test results to reward teachers with merit pay, on the assumption that teachers will respond to incentives for success, like participants in business organizations; the professional educators (or at least their unions) vigorously reject merit pay as a breach of professionalism that will undermine morale. The policymakers have endorsed transparent reporting of student performance, as opposed to norm-referenced reporting, so that parents can find out how their children are doing in relation to a meaningful standard. When the governing board of the National Assessment of Educational Progress authorized the creation of "achievement

levels” (basic, proficient, advanced) in 1990 to replace norm-referenced proficiency scores, prominent members of the research community objected to the change (and continue to object) on mainly technical grounds.

Policymakers enacted laws in nearly forty states to permit the creation of public charter schools, hoping that their freedom from excessive government regulation would encourage higher performance. Educators were skeptical and, in some cases, openly objected to what they saw as a diversion of public funds to quasi-public (or quasi-private) schools.

Policymakers have supported the use of contracting to allow private companies to manage schools. Educators have seen this move as a threat to public education and, in some cases, have openly fought against the awarding of contracts to for-profit companies.

Policymakers have pushed for use of school report cards so that parents can find out how their children’s schools are doing. They have also promoted the idea that failing schools would be subject to “reconstitution” (that is, reopening with a new staff and new principal), state intervention, and, in a worst-case scenario, a takeover of the school or school district by the state. None of these ideas emanated from educators, who continue to believe that the root problem of school failure is lack of resources.

The policymakers’ pressure for accountability has not run into a brick wall of resistance. It would be more accurate to say that it has plunged into a bowl of Jell-O™, in which demands for accountability are eventually but inevitably transformed into demands for more resources. Educators want to improve student performance, but to do so they must have higher salaries, smaller class sizes, more training, and so on. The starkest illustration of this transaction can be found in Massachusetts, which passed an ambitious school reform law in 1993. As part of the plan, the state pledged to put up an extra one billion dollars every year, on the understanding that students would be expected to pass state examinations

for high school graduation by 2003. The state did put up the money it promised, but by 2000, many educators were in revolt against the state testing program. The state's teacher union even ran an expensive advertising campaign and sponsored legislation to roll back the implementation of the state graduation tests.

Perhaps the most intriguing aspect of the debate over standards and accountability is that the states that have persisted in this strategy over time have seen steady improvement in student performance. North Carolina, Massachusetts, and Texas saw strong achievement gains for their students, both on state tests and on the regular tests administered by NAEP. The gains were especially significant for black and Hispanic students, whose performance in all states lagged far behind their white and Asian peers. In Virginia, which set high standards and aligned their state tests to those standards, the initial results were nothing short of appalling. But by the third year of testing, as teachers became familiar with the state's curriculum and as schools took accountability seriously, student performance was ratcheted up. In 1998, for example, the first year that the state tests were administered, only 40 percent of students passed the Algebra I test; by 2001, the proportion who passed had grown to 74 percent. Here, too, African American students made steady test-score gains.

As policymakers and educators jostled over the fate of accountability programs, another version of accountability lurked on the sidelines: vouchers. The theory of vouchers was that marketplace accountability would cure the ills of the schools; with sufficient information and the freedom to choose, went the theory, parents would withdraw their children from poor schools and send them to better schools. Public school leaders and teachers' unions railed against vouchers; they charged that they were unconstitutional because parents might use public funds to send their children to religious schools and warned that vouchers would destroy public education by allowing the most motivated families to flee to private and religious schools.

Vouchers directly challenge the supremacy of the state system of public education, so it is not surprising that spokesmen for public education would vigorously attack them. Vouchers are a form of accountability because they offer parents the opportunity to exit an institution that does not satisfy them. This is an alarming promise to professionals whose livelihood is dependent on the survival of that institution.

Thus far, the political battles over vouchers have limited the implementation of this policy initiative to only two cities (Milwaukee and Cleveland) and one state (Florida). The Democratic party is strongly allied to the teachers' unions, is completely opposed to vouchers in any form, and can barely countenance even public school choice. The Republican Party is ideologically sympathetic to vouchers because its preference for the free market predisposes it to embrace market solutions to social problems. But the Republican Party is irresolute about vouchers because the issue has a weak base within the Republican Party: rural districts don't care about school choice, as they ordinarily have only a few schools in their district; suburban districts are not animated about school choice because their students' performance is usually above the state's average; urban districts, where student performance and graduation rates are low, rarely vote Republican. So, even though there is strong support among young African Americans and Hispanic Americans for various forms of school choice, these groups vote reliably Democratic and do not exert any political pressure for a reform they prefer.

There will continue to be clashes between the policymakers who seek accountability and the educators who seek to deflect it. We can expect to see policymakers pumping more resources into education with the expectation that more inputs will eventually produce better outcomes for students. To some extent, this is a reasonable assumption: teacher salaries should be high enough to attract well-educated college graduates into the classroom; school facilities should be ample; school supplies should be adequate to students' needs;

and teachers should get additional education to stay abreast of improved methods and knowledge.

We can also expect that demands for improved performance will not abate. The public will continue to insist that students should be able to read, write, use mathematics, and be generally well prepared for further education or for technical jobs when they graduate high school. If large numbers of students continue to be poorly prepared, the public is likely to conclude either that a generation of school reform has failed or that the reforms to date have been too timid. If that should happen, then interest in accountability through market reforms—that is, vouchers—is likely to have greater public support than it has until now. Albert Shanker presciently recognized that the failure of standards-based reforms might pave the way for market-based reforms. His premature death, however, canceled out the one prominent voice among professional educators who was ready to lead a campaign in support of a strategy of standards, testing, and accountability.

In the near term, American education will continue to be driven by the two paradigms: the professional education paradigm, which deeply believes that the profession should be insulated from public pressure for accountability and which is deeply suspicious of the intervention of policymakers; and the policymaker paradigm, which insists that the public school system must be subject to the same incentives and sanctions based on its performance as are other large-scale organizations. How this conflict is resolved, and whether it is mooted by technological change in the delivery of education in the next decade or two, will determine the future of American education.