A Selective Social History of the Uses of Reading Tests

Patrick Shannon
Pennsylvania State University

Reading tests are modern.¹ In fact, they represent the liberal stance on education as few other artifacts can. Reading tests embody assumptions from the Enlightenment, capitalism, and cultural Darwinism; they project aphorisms such as science is progress, efficiency is paramount, and Anglo-Saxon tradition is virtuous. During their history, they have played many roles, but they have always served just one purpose—to hide the social construction of privilege behind a cloud of scientific objectivity. Since their invention just before the turn of the century to their recent invocation in the debates about the effectiveness of schooling in California, reading tests have been used to mold public education and to shape professional and popular perceptions of reading and reading education.

Ironically, during the last decade, support for reading tests among educational scientists has begun to wane. Tests were once considered the bottom line of reading programs, but their status has now sunk (Neill & Medina, 1989). For example, Richard Anderson, past president of the American Educational Research Association, wrote: "The strength of a standardized reading test is not that it can provide a deep assessment of reading proficiency, but rather that it can provide a fairly reliable, partial assessment cheaply and quickly" (R. Anderson, Heibert, Scott, & Wilkinson, 1985, p. 98). This valuation is quite a fall from the exalted heights that

¹Throughout this chapter I use the term reading tests to refer to norm-referenced standardized reading tests.
CHAPTER 5

William S. Gray set in 1915: “Standard reading tests supply information concerning all phases of instruction from broader issues involved in the course of study to the detailed difficulties encountered by individual pupils” (p. 141).

Having recently changed their minds, educational scientists have seemed confused by the continued, even expanded, use of standardized reading tests. Roger Farn (1992) tried to “solve the reading assessment puzzle” (p. 17). Jack Pikulski (1990) explored “the role of tests in a literacy assessment program” (p. 8). To educate test users and the public, these scientists engaged in campaigns to expose the problems of reading tests and to propose alternative “more scientific goals and means for reading assessment” (e.g., S. G. Farn et al., 1992). Despite their concerns, confusion, and campaigns, U. S. students have continued to take more and more reading tests: unit tests, district tests, state competency tests, and by 1999 a national reading examination at the end of fourth grade. Neill and Medina (1987) estimated that U.S. students take over 105 million standardized tests annually.

In what follows, I explore why the public, legislators, and many educators have clung to standardized reading tests while experts have begun to question their value. To address this question, I briefly examine the history of standardized reading tests. Reading tests are not just reified psychological principles; they are social artifacts produced in the continuous curricular debates about whose knowledge is most beneficial, who is considered literate and who is not, and what roles schooling and literacy programs play in the government’s and ruling class’s attempts at social engineering. Reading tests exist, are written in particular ways, and are used for social and political reasons as well as academic and scientific ones. To understand why the use and valuing of reading tests do not change as quickly as academic and scientific rationales is to understand that science is in the service of political ideologies, and not the reverse. Therefore, I investigate the social and political reasons for reading tests’ existence, development, and patterns of use.

Standardized reading tests were invented during the progressive era when science and business were thought to hold the answers for all social problems. During the decades at the turn of the 20th century, U. S. schools appeared unable to serve society and individuals as the United States changed from a nation primarily of small towns run according to face-to-face contact among citizens to one based on growing urban centers organized around industrial and commercial interests (Cremin, 1988). Changes in the social organization of the nation also altered traditional family life and values and further hindered children’s chances of making sense of the modern world. Moreover, the millions of immigrants settling in cities added

USING OF READING TESTS

foreign languages and customs to the new cultural challenges already facing public schools. Urbanization, industrialization, and immigration forced the public and educators to re-examine elementary and secondary school curricula to prepare citizens for the 20th century.

Then as now, although most agreed that change was necessary, there was much less agreement about what these changes should be (Kliebard, 1986). Traditionalists from William Torrey Harris at the turn of the century through E. D. Hirsch today favored continuation of European content and values over tests and technique. Child-centered advocates from G. Stanley Hall to David Elkind prized development and creativity over measurement and method. Social reconstructionists from Lester Frank Ward to Maxine Greene promoted access and equity over evaluation and erudition. Educational scientists from Joseph Mayer Rice to P. David Pearson argued that science and market would objectively remake schools and would equip both individuals and society for whatever changes might follow. Reading tests are one way to track how this last position became the dominant one in U. S. schools and social discourse about education.

DISENCHANTMENT OF THE MIND

Standardized tests of reading were first used to explore the processes of perception, thinking, and learning. They were instrumental in the creation of the academic discipline of psychology and its efforts to separate the study of the mind and its functions from the field of philosophy (Venezky, 1984). Discussions of Plato’s theories of remembrance, Rousseau’s theories of mental transactions, Froebel’s mysticism, and the like were considered too subjective to calculate the workings of the mind. Philosophical explanations of the mind required leaps of logical faith that could not necessarily be verified beyond argumentation. On the contrary, psychological investigations produced empirical data from which explanations of the mind could be built objectively:

The judgments of science are distinguished from other judgments by being more impartial, more objective, more precise, and more subject to verification by any competent observer and being made by those who by their nature and training should be better judges. Science knows or should know no favorites and cares for nothing in its conclusions but the truth. (Thorndike, 1906, p. 265)
From their inception, reading tests connoted objectivity, precision, and power. Standardization of design, implementation, and analysis were necessary to distinguish results from typical teacher judgments, to reduce error, and to make results comparable. Following John Locke's writings about the mind, psychologists theorized that mental activity was based on perception and sensation. They designed reading tests that would first help them to investigate sensation, mental speed, attention, recall, and the basic association of ideas through formalized activities involving decontextualized letters, syllables, words, and sentences as stimuli. From such data, they sought to build explanations of how perception led to understanding. To explain understanding, reading tests used paraphrasing (Huynh, 1908/1968), free recall (Romnes, 1884), and questioning (Thorndike, 1917) to glimpse what readers remembered or recognized from texts ranging in length from phrases to short paragraphs. This approach, of course, was in keeping with contemporary associationist theories and, later, behaviorist theories of reading: "When the mechanics of reading, if we may use that phrase to distinguish the process of reading from the process of understanding, are mastered, the whole attention may now be concentrated on the significance of the passage" (Judd, 1914, p. 368).

Absent from these tests and psychological considerations of reading were personal engagement, critical response, or individual construction of meaning because these obvious characteristics of sophisticated reading seemed at the time to require a spiritual re-enchantment of reading behavior. To continue to study these aspects would be unscientific, emotional, and incomparable. Results of such study could not lead to true statements about either the human mind or reading. As early as 1886, James Cattell acknowledged these concerns: "The conditions of the experiments place the subject in an abnormal condition, especially as to fatigue, attention, and practice" (Cattell, 1886, p. 63). Yet, the results of the early psychological investigations of reading using reading tests were substantial enough to lead Paul Klop (1908) to conclude that "remarkably little experimental information has been added to what Huynh knew in 1908, although some of the phenomena have now been measured more precisely" (p. xiv). Reading tests were recognized as the objective way to disenchant the mind of the spiritual notion that some things may be unexplainable, to define reading once and for all, and to determine who could read.

SCHOOL EFFICIENCY

Thorndike's faith in science carried over from psychology to school management. Reading and other tests provided measures by which administra-

USES OF READING TESTS
tors could make objective, rational decisions about what content and methods should be included in the school curriculum. This scientific management of schooling was patterned after the efficiency movement in business and industry, in which contribution to profit was the only acceptable rationale for introducing or maintaining any factor in production processes. Metaphorically, reading test scores became the profit in the production of literate citizens. The disenchantment of traditional, spiritual, or cultural explanations of mind was extended into school practices as teachers and students were considered factors in production equations. In this way, scientific management and reading tests were used to standardize and reduce the subjective human contribution to education.

Not all school personnel enthusiastically supported the scientific management of schooling, as Leonard Ayres (1915) reported in this retrospective:

Eighteen years ago the school superintendents of America, assembled in convention in Indianapolis, discussed the problems then foremost in educational thought and action. At that meeting a distinguished educator—the pioneer and pathfinder among the scientific study of education in America, Joseph Mayer Rice—brought up for discussion the results of his investigations of spelling among the children in the school systems of nineteen cities. These results showed that, taken all in all, the children who spent forty minutes a day for eight years in studying spelling did not spell any better than the children in the schools of other cities where they devoted only ten minutes per day to the study.

The presentation of these data threw that assemblage into consternation, dismay, and indignant protest. But the resulting storm of vigorously voiced opposition was directed against the investigator who had pretended to measure the results of teaching spelling by testing the ability of children to spell.

In terms of scathing denunciation the educators there present and the pedagogical experts, who reported the deliberations of the meeting in the educational press, characterized as silly, dangerous, and from every viewpoint, reprehensible, the attempt to test the efficiency of the teacher by finding out what the pupils could do. With striking unanimity, they voiced the conviction that any attempt to evaluate the teaching of spelling in terms of the ability of the pupils to spell was essentially impossible and based on a profound misconception of the function of education.

Last month in the city of Cincinnati, that same association of school superintendents again assembled in convention devoted fifty-seven addresses and discussions to tests and measurements of educational efficiency. The basal
proposition underlying this entire mass of discussion was that the effectiveness of the school, the method, and the teacher must be measured in terms of the results secured. (pp. 85–86)

By 1911, scientific management was so well accepted that the National Society for the Study of Education's Department of Superintendents established the Committee on the Economy of Time in Education to eliminate nonessentials from the elementary school curriculum, to set minimum standards for each school subject, and to improve teaching methods through the use of standardized tests. To gather appropriate data to accomplish these tasks and to measure compliance and performance later, ability tests had to be translated for large-scale implementation. "In the first place, it is undoubtedly desirable that the reading ability of students in one system of schools should be compared with the reading ability of children in other centers. For the purpose of such a general comparison, it is probably desirable that norm scores of standard tests be developed" (Judd, 1915, p. 562).

Reading tests became both the means and the ends of the scientific management of reading lessons and curricula. Advocates of scientific management began with two assumptions: that contemporary psychological definitions of reading were accurate and complete and that objective and precise instruments to measure that definition were possible and desirable. E. L. Thorndike put it this way in 1914:

It is obvious that educational science and educational practice alike need more objective, more accurate and more convenient measures of (1) a pupil's ability to pronounce words and sentences seen; (2) a pupil's ability to understand the meaning of words and sentences seen; (3) a pupil's ability to appreciate and enjoy what we roughly call "good literature"; and (4) a pupil's ability to read orally, clearly, and efficiently ... Any progress toward measuring how well a child can read with something of the objectivity, precision, co-measurability, and convenience which characterize our measurements of how tall he is, how much he can lift with his back, squeeze with his hand, or how acute his vision, would be a great help in grading, promoting, and testing the value of methods of teaching. (p. 1)

By 1919, W.S. Gray published Principles of Methods in Teaching Reading as Derived From Scientific Investigation, which included 48 tested principles covering skill knowledge norms for student progress throughout the grades, suggestions for oral and silent reading instruction, even specification for the

USES OF READING TESTS

printing of books to maximize the economy of reading. Using test score gains as the sole criterion for recommendation, Gray (1919) reported that "much reading of simple interesting material is effective in increasing the rate of reading" (p. 41), and "knowledge, while reading, that the material is to be reproduced improves the quality of the reading." He found that experiments had not shown one textbook method of teaching reading to be necessarily superior to all others regardless of circumstances; rather he maintained that instructional effectiveness varied according to how teachers used the materials, the backgrounds of the students, and the amount of materials available for instruction. Reading tests were not just to redesign the curriculum, they were supposed to redesign teaching, to produce verifiably literate students. "Tests will be most productive when individual teachers scrutinize their work carefully and record facts accurately, and reconstruct their methods on the basis of the facts secured" (Gray, 1918, p. 141).

Because of reading tests' newfound importance in managing schools, most leading reading experts and several educational psychologists produced their own versions of the tests (e.g., Buswell, Courtis, Dearborn, Gray, Judd, Thorndike, and later, Durrell, Gates, and Horn). Surveys of the reading tests began to appear as early as 1910 (see Bliss, 1918; C. Gray, 1917; W. Gray, 1916, 1917; Judd, 1914; Schmitz, 1914; Thorndike, 1914). Tests were available nationally by 1914. Competition for part of the reading test market was acute and served to reduce the variability among the reading tests being produced (Monroe, 1917). By the 1920s, reading tests became the primary tool for school improvement, program design, and judging student progress.

MEASURES OF CULTURE

Scientific management was expected to provide an objective means through which schools could be organized and run. Through the application of scientific methods, school personnel could justify the operations and outcomes to the public. Yet, the tools of scientific management in general and reading tests in particular were developed and designed by human beings, who held definite biases about society and all who inhabit it. Standardized tests encoded these biases, and an argument can be made that the tests themselves were developed and used to prove that these social biases were true. The following quotes trace these intentions from the originating premise during the Enlightenment, to the field of psychology in general, and to reading tests in particular:
CHAPTER 5

God has stamp'd certain Characters upon Men's Minds, which like their shapes, may perhaps be a little mended; but can hardly be totally alter'd and transformed into the contrary. He therefore, that is about Children, should well study their Natures and Antitudes, and see, by often trials, what turn they easily take, and what becomes them; observe what their Native Stock is, how it may be improved, and wit at it is fit for. He should consider, what they want; whether they be capable of having it wrought into them by industry, and incorporated there by Practice; and whether it be worth while to endeavor it. (Locke, 1705/1976, p. 67)

It may interest you to know the first [postwar] problem chosen for investigation by the Division of Psychology and Anthropology of the National Research Council is the problem of the mental and moral qualities of the different elements of the United States. What does this country get in the million or more Mexican immigrants of the last four years? What has it got from Italy, from Russia, from Scotland and Ireland? What are the descendants of the Puritans and Cavaliers and Huguenots and Dutch; and what are they doing for America? Psychology will undertake to do its share in an inventory of the human assets and liabilities of the United States, whenever it is asked to do so. (Thorndike, 1918, pp. 280–281)

Reading tests to the teacher cannot help but eventually mean not only concise and definite statements as to what she is expected to do in the course of study, but the reduction of instruction to those items which can be proved to be of importance in preparation for intelligent living and future usefulness in life. It will mean, too, an ultimate differentiation in training for the different types of children with which teachers now have to deal, and the specialization of work so as to enable teachers to obtain more satisfactory individual results. (E. P. Cubberley, 1917, p. xii)

Although standardized reading tests were supposed to eliminate teachers' bias from decision making about reading curriculum and instruction, in fact race, class, and gender biases were encoded in the theory, practices, and artifacts of the scientific management of reading through the use of these reading tests. Reading tests were to be the measure of success for some social groups whose "Native Stock," "mental and moral qualities," and "types" matched the expectations of psychologists and reading experts, and they also were used to justify the sorting of other social groups by their lack of the same factors to provide them "industry" and "practice" of an "ultimate" differentiation in training. From the beginning there were more measures of culture than true tests of literate ability. The systematic use of these tests meant success for the few, but failure for the many.

USES OF READING TESTS

These biases were not obvious in the tests themselves or in the rationales for testing. Whether individual or group, timed or untimed, reading tests employed the same formats to evaluate all readers. Because of the prevailing psychological theories, this practice seemed universally fair. Standardized directions, items, and passages, scoring procedures, and statistical scales attested to the egalitarian principles of testing. It soon became apparent, at least to some, however, that reading tests and the other standardized tests were stratifying students according to culture: "In several large cities, the school child, because of his unfavorable reactions to the school situation, comes in for clinical diagnosis. Since it is the child's reaction to the school situation which is at fault, it is well to test him along the line of the special abilities which he is expected to develop under the conditions of the school situation" (Schmitt, 1914, p. 152).

The same biases that teachers demonstrate in grading pupils are also apparent in standard reading tests. Students of different language and experiential backgrounds score lower. For example, Italian pupils are seriously handicapped. The sections of the city in which these pupils live are such that factors other than mere lack of English in the homes are probably to be recognized as contributing to the low rank of these pupils on tests. (Jordon, 1921, p. 111)

Even when these biases were identified, prevailing theories in psychology and education (which reflected society's faith in science, capitalism, and Anglo-Saxon traditions) found fault with cultures, and later individuals, rather than with science and scientific management. Once students were identified as "inferior," "disabled," or "deficient" (as Nila Banton Smith, 1936, labeled them in her history, American Reading Instruction), individual students were in line for more reading tests. "Pupils because they have had difficulties with a subject are in most need of the best possible teaching" (Gates, 1927, p. 19). To determine the best approach, reading experts began with the standard achievement test results and then applied a battery of specific subtests: standardized test records, test of attitudes and interests in reading, test of level of reading ability, test of abilities in comprehension and interpretation, test of oral and written recall of reading, test of abilities in the mechanics of reading, test of abilities related to study (see Durrell, 1939). By 1930, these batteries of tests could be constructed at any grade level with several alternative tests for each of the seven subsections of information required.

In the first book devoted entirely to remedial reading, C. T. Gray (1922) reported "a survey of the literature upon remedial measures shows that the development of methods and devices for this type of work has not kept pace
with the development of methods of diagnosis" (p. 365). Shortly thereafter basal readers began to fill this void with scripted lessons designed to boost test scores. For example in 1925, William P McCall and Leolah Mae Crabbs published a six-book series, *Standard Test Lessons in Reading*:

The *Standard Test Lessons in Reading* are offered to the teachers in our school (Lincoln School at Teachers College, Columbia University) with confidence that their use will give pupils a rate of speed and power to comprehend exceeding that yielded by ordinary methods of teaching silent reading...Every lesson is a test and every test is a lesson...Not only is every test a lesson, but every test is a standard test; that is, it shows how well the normal or typical pupil would read these same lessons. (p. xi)

These basalts furthered the efforts to manage reading lessons scientifically and to measure students' culture (Shannon, 1989). The teacher guidebooks, with scripted lessons and scope and sequences of skills, directed teachers to use the principles apparently derived from experiments. At the same time, the stories and the featured skills promoted what have come to be called Anglo-Saxon middle-class values and social habits (Luke, 1988). The basalts, then, provided the technical control that would enable teachers to work scientifically and culturally toward verifiably literate students capable of fulfilling their roles in society:

Within recent years various types of organizations of materials similar to published standardized tests have appeared...The better types represent an effort to develop a comprehensive series of exercises that include in well printed and illustrated forms the best informal devices and tests that the teacher would otherwise have to prepare herself. They are organized systematically to provide learning of the basal words, word recognition, comprehension, appreciation, and study techniques and simultaneously to test and diagnose ability and difficulty. Their purpose is to provide the nearest possible approach to the policy of daily diagnosis with the least expenditure of time of the pupils and teachers. (Gates, 1937, p. 374)

By the 1950s, surveys of school practices concluded that over 90% of teachers used basal materials on a daily basis and 98% of school districts administered standardized norm-referenced reading achievement tests at least once annually (Austin & Morrison, 1963; Barton & Wilder, 1964).

This interconnection between teaching and standardized tests became law, and large amounts of federal funding were directed to schools during the 1960s (House, 1978). Senator Robert Kennedy was apparently concerned that school administrators would divert federal funding for Title I reading programs away from people who were poor, unless a "reporting" requirement was added to the enabling legislation (McLaughlin, 1975). The purpose of the requirement was to inform parents so they would have access to the facts and figures to judge their children's progress. When this legislative requirement was operationalized through the federal government's Planning, Programming, and Budgeting Systems Office, the requirement was translated into scientific management terms in which standardized reading tests used to place students in the program would also become the means by which individual and program success would be measured and reported. Echoing to Judd's remarks of 1915 and presaging current opinions, Education Commissioner Keppel argued in 1964 that public reporting would provide performance comparisons among schools, districts, and even states and would thus stimulate competition. This renewed interest in scientific management was relabeled "system analysis," but reading tests were still considered profits to be counted.

The illusions of scientific objectivity—standardization, universality, and mathematical precision—obscured cultural, racial, and social class biases encoded in the formats, procedures, and topics of reading tests. The results from these tests became self-fulfilling prophecies that produced biased treatments of different social groups under the guise of scientific truth. In this way, reading tests not only define reading and measure school and programmatic success, but they verify White middle- and upper-class superiority.

**AN END TO HISTORY**

In light of this selective social history of reading tests, continued support for reading tests and the testing of reading makes sense, despite recent concerns among some psychologists and reading experts. Test advocates and users have simply been following the liberal traditions of psychology which have directed schooling and educators during most, if not all, of the 20th century. Although these traditions have prepared these groups to seek more scientific tests, more efficiency in reading instruction, and more affirmation of our national heritage, they do not prepare or enable any of us to call for the abolition of reading tests and the ending of the testing of reading.

Tests and testing are embedded in the fabric of schooling—its definition, its organization, its functioning, and its public legitimacy. They have been encoded in federal legislation, state laws, and district policy during the
CHAPTER 5

school lives of school personnel and taxpayers. In a sense, reading tests are part of our history both as students and as educators. To deny reading tests is to deny our schooling, our teaching, our literacy, even ourselves. It takes more than a few psychologists' and reading experts' contrary opinions to end the history of reading tests and the testing of reading. Reading educators, administrators, and the public are still told:

We must let scientific evidence answer questions about the reading process. (Stanovich, 1994, p. 280)

America will become a nation of readers when verified practices of the best teachers in the best schools can be introduced throughout the country. (Anderson et al., 1985, p. 120)

They often argue that the approach cannot be evaluated because it is expected to affect competencies other than those reflected by standardized measures. Fine, so identify those competencies, and competent reading researchers will find ways to measure them. (Pressley, 1994, p. 190)

Although nationalism may be regrettable in some of its worldwide political effects, a mastery of national culture is essential to mastery of the standard language in every modern nation... Children also need to understand elements of our literacy and mythic heritage, for example, Adam and Eve, Cain and Abel, Noah and the Flood, David and Goliath, The Twenty-third Psalm, Humpty Dumpty, Jack Sprat, Jack and Jill, Little Jack Horner, Cinderella, Jack and the Beanstalk, Mary Had a Little Lamb, Peter Pan, and Pinocchio. Also Achilles, Adonis, Aeneas, Agamemnon, Antigone, and Apollo as well as Robin Hood, Paul Bunyan, Satan, Sleeping Beauty, Sodom and Gomorrah, the Ten Commandments, and Tweedledum and Tweedledee. (Hirsch, 1987, p. 30)

We will develop American Achievement tests for 4th, 8th, and 12th graders in the five core subjects. (George Bush, national educational strategy speech, April 18, 1991)

I offer a social, not a technical, history of reading tests because it is impossible to separate human intention from the tests or the tests from the testing. Nor is it possible to separate the testing and testmakers from all the social pain that surrounds the continued scientific management of reading programs. As I have tried to demonstrate using advocates' own words, reading tests are designed to reduce reading to a psychological shell, to direct the teaching of reading, to sort social groups, and to protect the feelings of superiority of certain social groups over others. Regardless of the care given

USES OF READING TESTS

to these concerns in recent efforts to modify tests (e.g., Michigan and Illinois state examinations, the National Assessment of Educational Progress, etc.), these intentions remain in the tests and testing. Changing reading tests cannot change biased minds. Abolishing reading tests cannot abolish liberal ideology and history, which legitimate and perpetuate an unequal status quo in and out of schools. To end this history, we must wrestle schooling away from psychology and psychologists; we must deny reading education as a market to be fought over by multinational corporations; and we must stop reading education as cultural imperialism through both the forms and content we valorize. One lesson of this selective history is that we must invent postliberal reading education.