

Reconciling a Tradition of Testing with a New Learning Paradigm

Carole J. Gallagher^{1,2}

The entrenchment of standardized assessment in America's schools reflects its emergence from the dual traditions of democratic school reform and scientific measurement. Within distinct sociohistorical contexts, ambitious testing pioneers persuaded educators and policymakers to embrace the standardized testing movement. Their efforts ushered in an era of unprecedented testing in schools, a practice that has escalated despite challenges from contemporary psychologists who endorse a more learner-centered approach to education. This article examines the historical rationale for testing and urges educators to accommodate the legacy of standardized assessment within a new learning paradigm.

KEY WORDS: standardized tests; history; school assessment; educational testing; learner-centered principles.

Educators today face a dilemma. Should they support current presidential, legislative, and corporate initiatives that claim to ensure a quality education for all children through the escalation of standardized measurement of predetermined learning outcomes? Should they accommodate standardized testing within a contemporary learner-centered paradigm, which endorses a more eclectic "toolbox" approach to assessment that allows the informed educator to select among diverse gauges of learning progress?

Increasingly, parents, teachers, and school administrators seek advice from educational psychologists while weighing the social and academic consequences of standardized testing. Stakeholders ask critical questions about

¹Department of Counseling and Educational Psychology, Indiana University, Bloomington, Indiana.

²To whom correspondence should be addressed at 1030 Champion Ct., Columbus, Indiana 47201. E-mail: cjgallag@indiana.edu

appropriate uses for standardized test scores, and express a need to understand the rationale for the escalation of standardized assessment. Why do American schools continue to rely on group-administered, standardized test scores for educational decision-making purposes? How did this powerful historical tradition become a foundation for educational practices, and how might educators accommodate standardized testing with learner-centered principles [American Psychological Association (APA), 1993]?

This article seeks to answer these questions by examining the unique social and historical circumstances that converged to institutionalize an educational movement that has profoundly impacted America's schools. It acknowledges the vision of testing forefathers who (a) championed the democratic ideal of ensuring that all American schoolchildren would have access to opportunities for educational advancement, and (b) imagined that progress in developing increasingly sophisticated measurement tools would result in a more objective and equitable system of assessment. But it also suggests that educators must engage in deliberate strategies to reconcile the legacy of standardized testing within a learner-centered framework that has emerged in response to the intervening socioeconomic forces of this century (Chapman, 1988; Hanson, 1993; Heubert and Hauser, 1999; Jennings, 1998; Popham, 1981; Sacks, 1999; Thorndike and Lohman, 1990). It challenges twenty-first-century educators to chart a new course for the trajectory of assessment in America's schools.

TRACING THE ORIGINS OF TRADITION: HORACE MANN'S COMMON TEST

By the mid-nineteenth century, it was clear to American philosophers, scientists, and educators that the popular college tradition of oral qualifying examinations was flawed:

Public examinations were generally held once a year and were more in the nature of public displays or exhibitions to show off brilliant pupils or to glorify teachers. It was as a result of abuses to which such displays gave rise and of the criticisms which they prompted that written exams began to be introduced. (Kandel, 1936, p. 24)

It was within this era of discontent that Horace Mann introduced his vision for reforming American education through common schools (Linden and Linden, 1968). In 1845, he proposed that the schools become vehicles for social advancement, enabling all citizens to live educated lives. Mann persuaded the Boston Public School Committee to allow him to administer written exams to the city's children in place of the traditional oral exams. Using a common exam, he hoped to provide objective information about the quality of teaching and learning in urban schools, monitor

the quality of instruction, and compare schools and teachers within each school.

The test results indicated that there were wide gaps in the knowledge of Boston's schoolchildren, and Mann's proposals for additional testing were heeded in hopes of securing a fail-safe method for determining which students were prepared to move to the next academic level. His model was so successful that competitive written exams were adopted by school systems in nearly all U.S. cities, and in 1865, the New York Regents Exams were developed on the basis of Mann's assessment concepts. Ralph Tyler (in Houts, 1977), a noted curriculum historian, concluded,

At a time when the need for universal education was developed, the testing movement furnished both an ideological and an instrumental basis for the practice of schools and colleges in sorting students rather than educating them . . . it promoted the simplistic notion that important outcomes of schooling could be adequately appraised by achievement tests. (p. 17)

INITIAL DRIVING FORCES

In 1859, Charles Darwin published his remarkable findings, highlighting individual differences as being fundamentally important to the future of the human species. European psychologists were studying individual mental differences and deficiencies, creating laboratory experiments, and attempting to measure learning and human behavior scientifically. An influx of immigrants, compulsory school attendance mandates, restrictive child labor laws, and a rising cost of living created strong motives for scientists to construct more effective and efficient mechanisms with which to sort students in America's crowded classrooms.

The late nineteenth century also witnessed the popularization of the elective curriculum. Increasingly, diverse coursework and subjective teacher judgments made the task of student assessment more complex. The necessity to gauge progress in each discipline resulted in a proliferation of independent tests of achievement, despite an awareness that written scores could also be misleading in that "instructors may attempt to influence the process in their favor by teaching to the test" (Hanson, 1993, p. 199). But the convenience of using objective tests outweighed this skepticism, and school administrators began systematically collecting data to construct comprehensive and comparable portraits of student learning.

THORNDIKE'S PIVOTAL ROLE: SCIENTIFIC MEASUREMENT

In 1898, a Columbia University professor E. L. Thorndike was experimenting with objective tests, quantifiable scales, and efficient surveys.

Believing that previous forms of assessment reflected subjective opinion, he spearheaded a campaign for more scientific measurement tools. Thorndike believed that society would benefit from systematic identification and segregation of students according to their intellectual abilities (Hanson, 1993). Envisioning a utopian society profiting from optimal use of human resources, he applied scientific principles to education. He said,

Educational agencies are a great system of means not only of making men good and intelligent and efficient but also of picking out and labeling those who for any reason are good and intelligent and efficient . . . They help society by providing it not with better men but with the knowledge of which men are good. (Thorndike, 1913)

The esteemed Dr Thorndike's recommendations contained professional clout, and from 1900 to 1910, urban schools in Pennsylvania, New Jersey, New York, Massachusetts, Michigan, Kansas, and California began using the new measurement tools (Chapman, 1988).

BINET'S TESTS OF INTELLIGENCE

Also during the first decade of the twentieth century, a French physiologist/psychologist was making extraordinary contributions to mental testing. Alfred Binet developed an individually administered test of intelligence for use in identifying learning deficiencies in "slow children who would not profit significantly from schooling" (Walsh and Betz, 1995, p. 2). Binet created an intelligence scale that identified the mental age of children, and test administrators could then graph and compare individual scores on this scale.

Inspired by Binet's strategy for eliminating retardation from school systems, H. H. Goddard brought Binet's model for testing intelligence to the United States in 1911. Goddard's mission was to convince public school officials to incorporate student intelligence testing into their decision-making processes. One year later, William Stern proposed the present IQ formula, that of dividing the mental age by the chronological age and multiplying by 100. Support for this dynamic movement was grounded in the positivistic assumptions that dominated psychological thought; it defined the testing of mental capacity as scientific by establishing guidelines for its systematic measurement.

EXPANSION INTO SCHOOLS: TERMAN'S MISSION

In 1916, Stanford's Lewis Terman interpreted, revised, and expanded the IQ test and renamed it the Stanford-Binet Test of Intelligence. Although it was initially designed as a practical tool to identify those who were

“feble-minded,” Terman realized its potential usefulness within schools for facilitating educational placement and career tracking (Terman, 1916).

Social scientists embraced the use of technical devices for sorting people according to ability, or according to who was fit and capable. They believed that social prosperity hinged on critically assessing human intelligence and ranking people according to their capacities. Educators assumed that scientific evidence now existed to explain individual differences and concluded that poor performance must reflect an inherent lack of ability. According to Gould (1981, p. 195), “A technology had been developed for testing all pupils. Tests could now rank and stream everybody; the era of mass testing had begun.”

ADDRESSING WAR PRIORITIES: ARMY ALPHA

Spurred by the urgencies of impending war, the U.S. Army sought to identify likely officer candidates from the huge pool of recruits. Their aim was to place recruits expeditiously in positions where they would be most productive. Arthur Otis and Robert Yerkes, who had been developing tests of group intelligence, agreed to create a prototype paper-and-pencil, multiple-choice test to measure soldiers' mental abilities, a test that became known as the Army Alpha Test. This format was envisioned as the most effective way to test large groups of people, and it became the model for all subsequent standardized tests (Rothman, 1995). Otis and Yerkes also devised a means of objectively scoring these tests so results could be recorded and disseminated efficiently.

Nearly two million men were examined, sorted, assigned, trained, and discharged using the Army Alpha and Beta tests during World War I. According to Hanson (1993, p. 212), “The war changed the image of tests and of the tested . . . they [the tests] were now legitimate means of making decisions about the achievements and aptitudes of normal people.” Copyright offices were overwhelmed by requests for patents on standardized intelligence tests. Psychologists, schools, hospitals, and social agencies recognized the potential of these assessment tools for determining levels of mental functioning and facilitating comparisons among examinees.

In 1919, Terman transformed the Army Alpha into the National Intelligence Tests for schoolchildren, and over 400,000 copies were sold during the next 11 months (Hanson, 1993; Terman, 1919). Although many psychologists continued to develop tests of mental ability, the assessment of mental functioning was not limited only to intelligence. A wide variety of human traits was now measurable, and tests began to legitimize the technical division of labor by sorting personnel to fill specialized positions.

THE SCHOOLS JUMP ABOARD

After witnessing the wholesale success of the Army testing program and its ready application in clinical and government settings, educators at the K-12 and college levels sought new and better ways to predict, diagnose, and explain learning differences. They hoped to capitalize on the military's expertise and adopted instructional models that resembled efficient military grouping procedures. In this era, academic tracking became entrenched in schools as scores from intelligence tests dramatically changed the ways in which students were classified. Standardized tests were used to stratify students of different abilities into different curricular paths, thereby restricting their academic and social choices (Zanderland, 1998).

Social factors such as booming urban enrollments, massive school-building programs, and the need for industrial efficiency increasingly pressured schools into describing student performance in terms of test results. Subsequently, educational researchers called for a more diversified school curriculum, one that provided a different education for different students and was more suited to their assessed abilities.

In response, the Stanford Achievement Tests were published in 1923, as a battery of tests for elementary students that combined several content-area tests into one exam. By 1929, more than five million tests were administered annually, and results were used to segregate those who had learned from those who had not (Thorndike and Bregman, 1934). Standardized test scores also documented instructional effectiveness. Ironically, the same principles applied to sorting students were now being used to sort the schools themselves.

In 1929, the University of Iowa created the first sets of student achievement tests to be administered statewide on a voluntary basis: the Iowa Test of Basic Skills and the Iowa Test of Educational Development. Eager to embrace the cost-efficient school assessment initiative, other states began using Iowa Tests, and for over 50 years they remained the most frequently used commercially available achievement tests in the nation (Peterson, 1983).

THE NEED FOR TESTS OF APTITUDE

Though enjoying widespread application, tests of intelligence and achievement nonetheless had limitations. Although they could provide valuable information about global intelligence and past learning, they revealed little about specific abilities or predicted performance. Such assessments, or aptitude tests, were touted as a means for estimating or predicting a person's ability to learn or perform if given the opportunity. Social Darwinists held

that people could be adroitly channeled toward their potentials, and human resources would be used more efficiently.

College admissions personnel, for whom the identification and recruitment of qualified students was a top priority, sought to streamline the college admissions process through the adoption of standardized college entrance requirements. In 1923, a consortium of college officials, the College Entrance Examination Board (CEEB), was impaneled to convert a set of common admissions standards into an examination that would serve as an admissions criterion for all member colleges. The CEEB agreed to oversee the administration of this combination achievement/intelligence test to students seeking admission to member colleges.

Two years later, Carl Brigham of Princeton refined the CEEB test, which became known as the Scholastic Aptitude Test (SAT); thereafter, it would help define the nature and content of college preparatory instruction (Walsh and Betz, 1995). Brigham voiced a concern shared by colleagues: ensuring equal educational opportunity for all students. In 1936 (p. 4), I. L. Kandel of the Carnegie Foundation concurred, recommending more comprehensive assessment to provide a portrait of examinees that facilitated the determination of "the right education for the right individual."

In 1947, the Educational Testing Service was established as a nonprofit agency to oversee CEEB. The National Merit Qualifying Tests and its scholarship programs were added to the assessment repertoire in 1957. The American College Test (ACT), created in 1959 as an alternative to the SAT, became widely accepted for college admissions and financial aid decisions (Walsh and Betz, 1995). Both tests have endured revision processes since that time and present annual documentation of internal validity and reliability. The ACT was updated in 1989 to stress abstract thinking skills, and the SAT's revisions within the past 10 years have resulted in recentering, the addition of subject level tests (SAT II), and a name change (Scholastic Assessment Test).

STANDARDIZATION: A MEANS OF ENSURING EQUITY

America's efficient ranking and sorting of students necessitated the institutionalization of a standard means of administering tests of aptitude and achievement (Hanson, 1993). Test publishers were obliged to document statistics about trial performances before releasing tests for use in schools. A standardized protocol was created to guide the administration of all large-scale tests. Rigid instructions called for the isolation of a group of testees, the precise following of standardized administrative instructions, use of a nationally endorsed set of multiple-choice questions about a broad survey of skills and knowledge, and external scoring of tests (Hanson, 1993). Student scores could then be compared on an intraschool and interschool basis.

Proponents insisted upon carefully monitored testing conditions as a means of promoting score reliability, assessment fairness, and optimal student performance. As Goodenough (1949, p. 87) said, "The early tests were far from perfect; nevertheless, they called attention to the necessity of using a standard situation, of providing a common basis of reference, if individuals are to be classified in a uniform and meaningful way." Accompanied by technological advances, these tests could be scored electronically, which was believed to increase scoring accuracy and efficiency.

COLD WAR AMPLIFICATIONS

Increasing industrialization and bureaucratization, flourishing capitalism, a second world war, economic depression, and a behaviorist psychological paradigm conspired to create a climate of academic urgency within America's schools. Piqued by Cold War tensions, American citizens and leaders had a heightened awareness that maintaining its competitive position in the world was dependent on identifying student talent in academics, leadership, and managerial skills (Wigdor and Garner, 1982). Standardized tests were used increasingly to determine which students should be promoted or retained, assigned to remedial or special education placements, and receive academic honors. Performance on a battery of tests could be used to determine the course of a student's academic or vocational future.

In 1965, the Elementary and Secondary Education Act (Title 1) required schools to administer standardized tests and submit their results to qualify for federal funds in subsequent years (Chapman, 1988; Thorndike and Lohman, 1990). This initiated the large-scale use of test scores for evaluating instructional programs systematically; by comparing individual and group test scores, educators and policymakers could make recommendations about the demonstrated effectiveness of specific methods of instruction.

In 1969, the federal government supported testing initiatives by expanding the National Assessment of Educational Progress (NAEP), which tested samples of students from various states in all subject areas to gauge national achievement. Later nicknamed the "Nation's Report Card" because it reflected student performance on tests administered in nearly every state, it compared district and state performances, and determined a national score that was used for international comparisons (Berliner and Biddle, 1995; Gould, 1981).

A FAIR TEST FOR ALL?

During the mid-1960s, the civil rights movement created a heightened awareness of potential testing inequities (Sacks, 1999). Critics argued that

standardized tests were biased in terms of social class and racial/cultural background, thereby reinforcing social and economic inequality. Although civil rights activists believed that testing inequities might be alleviated by suspension of testing practices, they were confronted with a dilemma; if they sought to eliminate state accountability programs that allegedly favored White, middle-class schools, they risked losing evidence of inadequacies in predominantly minority schools (Berliner and Biddle, 1995).

The debate over the fairness of testing America's students was passionate, and in 1966, the National Center for Education Statistics commissioned a study to examine issues of equity among racially and ethnically diverse student populations. Identified as the Coleman Report, one finding of this study revealed that the most important predictor of school achievement was the student's "general social context," or home background and related neighborhood factors (Berliner and Biddle, 1995). Critics of publicly funded education suggested that this finding implied that schools had little impact on academic performance because educational experience could not compensate for lack of ability; intelligence was not modifiable. Testing proponents then claimed that this also demonstrated that home environments, and not biases inherent in standardized assessments, were responsible for lower test scores in certain populations. Although these claims were later shown to be erroneous because of design and data analysis flaws, for many years the Coleman Report remained an indictment of school mediocrity and a testimony to the equity in standardized testing (Berliner and Biddle, 1995).

In 1978, Jane Mercer introduced the System of Multicultural Pluralistic Assessment in response to continuing allegations that existing standardized tests bore inherent racial biases that discriminated against minority students (Kincheloe et al., 1996). In particular, critics of standardized testing voiced concern that women, limited English-proficiency students, and racial minorities were penalized because of cultural differences. Because tests "are artifacts of culture, and culture may not diffuse equally into all households . . . subcultures vary in ways that inevitably affect test scores" (Kincheloe et al., 1996, p. 32). Assumptions about a universal body of "common knowledge" were challenged, igniting a controversy about inherent biases in intelligence and achievement testing that continues to be debated today (Sacks, 1999; Wigdor and Garner, 1982; Zanderland, 1998).

TESTING IN THE ERA OF ACCOUNTABILITY

The reevaluation of standardized test contents and design, testing procedures, and test uses and misuses by educators was not supported by businesses and government, who sought to extend the standardized school

experience to all American students. It was expected that schools be administered like corporations, with an unambiguous bottom line; test scores were concrete, reliable indicators that assured the attainment of minimum competencies by raising student achievement (Walsh and Betz, 1995).

The 1970s witnessed an escalation in the demand for educational accountability. For schools, test scores became essential weapons in defending against the loss of students, programs, or funding (Sacks, 1999). State governments (notably Michigan) considered authorizing the release of incentive monies to districts whose test scores significantly improved and the federal government offered grants to schools seeking to upgrade their assessment programs (Sacks, 1999). When scores declined, teachers and administrators were held accountable for rectifying the problem, which often meant implementing back-to-the-basics curricular changes to ensure a better match with test contents. Aggregate performances on tests such as SAT, primarily taken only by students seeking college admission, were often used to judge overall school, district, and state academic performance (Berliner and Biddle, 1995).

In 1974, Congress changed the structure of Title 1 testing and recommended expanded standardized assessment in schools for program improvement purposes. Thereafter, progress toward goals was measured (and schools funded) using standardized scores. By the 1980s, 33 states mandated some form of minimum competency testing and over 200 million tests were administered annually to determine IQ and academic readiness (Rothman, 1995).

A NATION AT RISK?

The release of *A Nation at Risk* (National Commission on Excellence in Education, 1983) fueled the testing frenzy with this ominous warning:

Our nation is at risk. Our once unchallenged preeminence in commerce, industry, science, and technological innovation is being overtaken by competitors from throughout the world. . . . the educational foundations of our society are presently being eroded by a rising tide of mediocrity that threatens our very future as a Nation and a people. (p. 5)

Schools, colleges, and universities were urged to adopt more rigorous standards and higher expectations for student performance. State-supported standardized tests to be administered at key transition points in schooling were recommended as appropriate measures for getting back on track (Rothman, 1995). Terrel Bell, then President Reagan's secretary of education, maintained a wall chart outside his office that graphed the progress of each of the 50 states toward meeting the nation's educational challenges.

The statistics cited in *A Nation at Risk* did not go unchallenged. Berliner and Biddle (1995) support critics who claimed that in 1985, students administered either the Iowa Tests of Basic Skills or California Achievement Test actually demonstrated substantial gains. Because such tests are recalibrated in 7-year cycles, to ensure that the typical student again scores near the 15th percentile, test designers anticipate initial overall score decline because recent achievement gains are erased. Furthermore, evidence mounted that despite the Commission's assertions about test score declines, standardized scores in verbal achievement had changed little in the previous decade, and math achievement scores showed modest improvement (Berliner and Biddle, 1995).

Nevertheless, by 1989, 47 states had responded to the report's recommendations by adopting policies that expanded statewide testing programs. Many local districts implemented their own plans to raise student scores by allocating more financial resources to testing budgets and by aligning curriculum with expressed testing purposes (Rothman, 1995).

A GENERATION OF HIGH STAKES TESTS

In 1991, a new national report card monitored state-by-state progress toward six national education goals defined by then President George Bush and his advisory board of state governors. Because nagging questions still existed about the quality of education, they promoted the development of more sophisticated standardized tests that more accurately and equitably portrayed student achievement. Local and national press encouraged parents to become informed educational consumers by attending to a school's publicized test scores.

In Clinton's 1994 Goals 2000: Educate America Act, nationwide administration of standardized tests remained a primary function of elementary and secondary schools. National goals were used to garner support for certification of "voluntary" national education and skill standards (Heubert and Hauser, 1999). Educators who supported this program believed it clarified what was expected of teachers and students on standardized tests and what instructional strategies might contribute to higher achievement scores. More than 35 states mandated test-based graduation requirements, acknowledging their responsibilities as gatekeepers (Heubert and Hauser, 1999). Educators in these states administered standardized assessments in core areas and routinely reported those results to parents, local boards of school trustees, and state departments of education.

A study conducted in 1994 by the National Center for Education Statistics revealed somber findings: average test scores for eighth graders in

high-achieving states (Iowa, North Dakota, and Minnesota) were comparable to the highest-achieving foreign countries (Taiwan and Korea); however poorly achieving states (Louisiana, Mississippi) ranked with the developing countries (Jordan). Although this range of scores undermined claims about "average" American achievement, it documented the potential for standardized testing to expose inequalities in educational opportunity.

FEDERALLY MANDATED ASSESSMENT

Toward the close of the twentieth century, American educators recognized that mandating standardized assessment and administering more competitive tests represented only one step toward improving student performance (ETS website, 2002; FairTest website, 2001). Spurred by videotaped analyses of classroom practices in Japan that demonstrated startling differences in teacher effectiveness, educators were anxious to explore instructional reform (Sacks, 1999). Yet the same study that yielded such compelling data about classroom activities (1998 Third International Math and Science Study) also contributed to bipartisan legislative support for President George W. Bush's Elementary and Secondary Education Act (ESEA, or "No Child Left Behind Act"). Riley, Clinton's Secretary of Education, insisted that the way to improve student achievement was to institute nationwide assessment of 4th, 8th, and 12th graders. His advice was heeded by both Republican and Democratic policymakers and transformed into a revitalized commitment to Bush's demand for accountability through annual testing of 3rd through 10th graders.

Title I of ESEA reveals ambitious plans to reform education, especially for disadvantaged children. By 2005, content and performance standards will be assessed in reading and math, and by 2007, science will be included. By 2014, these assessments will evaluate statewide progress toward nationally established proficiency goals and verify NAEP results (United States Congress, 2001). States must also implement an accountability system, with attached incentives, to ensure that all districts are demonstrating adequate yearly progress in achievement. Low-performing schools may face "corrective action," such as mandatory tutoring, replacement of school staff, or school restructuring (United States Congress, 2001).

The new law calls for academic assessment (though not necessarily standardized tests) that is valid and reliable for the purposes for which they are intended. It supports multiple measures of academic achievement and higher-order thinking skills, and recommends testing reports that include interpretive, descriptive, and diagnostic information (United States Congress, 2001). Although all states except Iowa currently use some form of

state-mandated assessment of standards, many states hope to continue to administer commercially available norm-referenced tests in lieu of constructing new measures (FairTest website, 2001). To date, only 15 states assess math and reading in the required grades, using standards-based tests. Limited funding has been authorized to support states in the local development and administration of innovative assessment standards and tools.

This law has encountered resistance from reformers whose vision for school improvement is focused on raising the quality of learning through multiple assessment methods such as classroom-based tests, essays, observations, projects, performances, and portfolios (FairTest website, 2001). They advocate raising student performance by addressing (1) the curricular needs of individual students within unique classroom contexts and (2) the quality of teaching that mediates such learning (APA, 1995). Although legislators also endorse these learner-centered goals, they believe the promises of a federally mandated testing program outweigh its disadvantages (ETS website, 2002). Educators face a future in which standardized testing will surely assert a dominant role.

THE TWENTY-FIRST CENTURY: FACING NEW CHALLENGES WITH TRADITIONAL TOOLS

The trajectory of the U.S. economy in the information age continues to be nudged away from basic manufacturing and toward consumer services and technology. Educational attainment and demonstration of technical skills are associated with power and wealth, and schools continue to be held accountable for producing students with the expertise to thrive in this competitive environment (Sacks, 1999). High test scores are presented as unbiased evidence that the student possesses the requisite cognitive competencies for academic achievement, and those who perform well on tests are granted access to opportunities for economic advancement (Berliner and Biddle, 1995).

Understandably, the high stakes nature of such tests produces anxiety in students, parents, and educators; a test score remains a valued piece of information about achievement to be considered during decision-making processes. School officials now routinely refer to test scores when determining who qualifies for graduation, promotion, and retention (Heubert and Hauser, 1999). Scores facilitate educational placement and ability grouping, steer curricular choices, and establish criteria for college and professional schools admission (Sacks, 1999). Test results are evaluated before awarding student financial aid packages, athletic scholarships, and merit awards, and consulted before finalizing occupational decisions. They impact the

allocation of state incentive monies to schools and are presented as justification for school choice and for the funding of education through vouchers.

Students, parents, and educators now face a dilemma. While historical review presents a compelling rationale for standardized testing, many legitimate concerns remain about standardized assessment (Beliner and Biddle, 1995; Chapman, 1988; Sacks, 1999). In fact, in the 2000 report from the NSSE (National Society for the Study of Education), educators suggest that the achievement testing movement as a whole has been a disappointment, having never fulfilled its potential to improve schools in the ways envisioned by its forefathers.

A NEW ASSESSMENT PARADIGM: THE LEARNER-CENTERED PRINCIPLES

In response, psychologists have attempted to reconcile their ambivalence toward the tradition of testing by providing educators with a framework for addressing lingering concerns about individual test performance without sacrificing instructional integrity. They endorse a shift that portrays learners as engaged, reflective learning transformers embedded within diverse learning communities. This profile is based on advances in brain-based neurocognitive and human development research, refinements in sociocultural theory, and the convergence of human information processing and a constructivist learning paradigm (Alexander and Murphy, 1994). Each has contributed uniquely to the reassessment of goals for all learners.

Since 1993, the APA has recommended that educators address the challenges presented by a changing socioeconomic climate through the adoption of a set of learner-centered principles. American Psychological Association anticipated that these 12 (expanded to 14 in 1995) interrelated recommendations would provide guidance to those seeking to identify learning strategies for educating American students within a culture of standardized assessment. Research supports the theoretical connection between these principles and the potential for meaningful measures of learning outcomes (Alexander and Murphy, 1994).

These principles focus on learner-controlled psychological factors that interact with environmental supports and constraints (APA, 1995). Based on theories about the nature of learning processes and social contexts, the construction of knowledge, motivation, and the development of learning and metacognitive strategies, these principles suggest that certain qualities may predispose the twenty-first-century learner to success: cognitive flexibility, self-direction, cooperation, resourcefulness, perspective-taking, the

ability to communicate clearly, and use of strategies such as planning and goal-setting.

The 14th principle specifically stipulates the need to set appropriately high, challenging standards for assessing learner progress and defining learning outcomes. American Psychological Association suggests that learners may be assessed (or self-assessed) at negotiable intervals, with feedback provided to them using a smorgasbord of evaluation strategies. Such assessments may include measuring problem-solving skills through evaluation of non-timed, real-world tasks (Berliner and Biddle, 1995). Information gathered through such authentic performance assessment, which is contextually aligned with individual or group problems in a natural setting, becomes a valid portfolio supplement to data about basic skills derived from individual performance on an achievement test. Student portfolios might exhibit traditional norm-referenced test scores as well as locally evaluated, criterion-referenced test scores. Such pluralism in assessment reflects a trend evident in American educational research for the past 20 years (Eisner, 2002).

Test scores may be used to reveal clues about a student's metacognitive awareness of formulating critical questions, weighing alternative responses, and generating creative solutions. Eisner (2002) explains,

The challenge to assessment is to somehow create tasks that give students opportunities to display their understanding of the vital and connected features of the ideas, concepts, and images they have explored. In short, the aim is to help students demonstrate that they have grasped ideas as part of a larger field and as historically situated elements within a community of discourse. (p. 205)

Exploring alternative pathways to the same outcome are encouraged, and multiple possible solutions are valued. Through knowledge transformation, students may demonstrate capacities to extend their learning to new domains; creative assessment will capture their ability to transfer meaningful content from one context to another. Educators may come to view such interpretive measurement tools as the complement, and not the antithesis, of standardized testing.

PUTTING A CENTURY OF TESTING IN PERSPECTIVE

In summary, historical analysis supports a rationale for standardized testing in America's schools. Yet testing has evolved since its conception as a common denominator within a society focused on scientific advancement, and a critical *reassessment* of measurement tools and their impact on all stakeholders ensures that the next generation of standardized assessment is rededicated to the democratic ideals for which testing was originally designed. By broadening the assessment focus to include collaborative

activities that foster the apprentice learner's skill development within a school community, educators may be effectively preparing students to approach standardized testing as a learning experience that only taps limited facets of their developing expertise.

ACKNOWLEDGMENT

The author thanks Dr David Flinders for editorial assistance with this article.

REFERENCES

- Alexander, P., and Murphy, P. (1994). *The Research Base for APA's Learner-Centered Psychological Principles*, Paper presented at the annual meeting of American Educational Research Association, New Orleans, April 4-8, 1994.
- American Psychological Association (1993). *Learner-Centered Psychological Principles: Guidelines for School Redesign and Reform*, American Psychological Association, Washington, DC.
- American Psychological Association (1995). *The Learner-Centered Psychological Principles: A Framework for School Design and Reform*, American Psychological Association, Washington, DC.
- Berliner, D., and Biddle, B. (1995). *The Manufactured Crisis: Myths, Fraud, and the Attack on America's Public Schools*, Addison-Wesley, Reading, MA.
- Chapman, P. (1988). *Schools as Sorters*, New York University Press, New York.
- Darwin, C. (1859). *On the Origin of Species by Means of Natural Selection*, Murray, London.
- Educational Testing Service (2002). Website: www.ets.org.
- Eisner, E. (2002). *The Educational Imagination: The Design and Evaluation of School Programs*, Pearson Education, Upper Saddle River, NJ.
- Goodenough, F. (1949). *Mental Testing*, Rinehart, New York.
- Gould, S. (1981). *Mismeasure of Man*, W. W. Norton, New York.
- Hanson, F. A. (1993). *Testing, Testing: Social Consequences of the Examined Life*, University of California Press, Berkeley.
- Heubert, J., and Hauser, R. (eds.) (1999). *High Stakes: Testing for Tracking, Promotion, and Graduation*, National Academic Press, Washington, DC.
- Houts, P. (ed.) (1977). *The Myth of Measurability*, Hart, New York.
- Jennings, J. (1998). *Why National Standards and Tests?* Sage, Thousand Oaks, CA.
- Kandel, I. L. (1936). Examinations and their substitutes in the United States. New York: Carnegie Foundation for the Advancement of Teaching.
- Kincheloe, J., Steinberg, S., and Gresson, A. (1996). *Measured Lies: The Bell Curve Examined*, St. Martins, New York.
- Linden, K., and Linden, J. (1968). *Modern Mental Measurement: A Historical Perspective*, Houghton Mifflin, Boston.
- National Center for Education Statistics (1994). *The Condition of Education, 1994*, Office of Educational Research and Improvement, U.S. Department of Education, Washington, DC.
- National Commission on Excellence in Education (1983) *A Nation at Risk: The Imperative for Education Reform*, U.S. Government Printing Office, Washington, DC.
- National Organization for Fair and Open Testing (2001). Website: www.fairtest.org.
- National Society for the Study of Education (2000). *99th Yearbook: American Education: Yesterday, Today, and Tomorrow*, University of Chicago Press, Chicago.
- Peterson, J. (1983). *The Iowa Testing Programs*, University of Iowa Press, Iowa City.
- Popham, W. J. (1981). *Modern Educational Measurement*, Prentice-Hall, Englewood Cliffs, NJ.

- Rothman, R. (1995). *Tests of Significance*, Jossey-Bass, San Francisco.
- Sacks, P. (1999). *Standardized Minds*, Perseus Books, Cambridge, MA.
- Terman, L. (1916). *The Measurement of Intelligence*, Riverside, Cambridge, MA.
- Terman, L. (1919). *The Intelligence of School Children*, Riverside, Cambridge, MA.
- Thorndike, E. L. (1913). *Introduction to Theory of Mental and Social Measurement*, Columbia University Press, New York.
- Thorndike, E. L., and Bregman, E. O. (1934). *Measurement of Intelligence*, Columbia University Press, New York.
- Thorndike, R., and Lohman, D. (1990). *A Century of Ability Testing*, Riverside, Chicago.
- United States Congress (2001). Website: www.congress.gov.
- Walsh, W. B., and Betz, N. (1995). *Tests and Assessment*, Prentice-Hall, Englewood Cliff, NJ.
- Wigdor, A., and Garner, W. (1982). *Ability Testing: Uses, Consequences, and Controversies*, National Academy Press, Washington, DC.
- Zanderland, L. (1998). *Measuring Minds*, Cambridge University Press, Cambridge, UK.

Copyright of Educational Psychology Review is the property of Springer Science & Business Media B.V. and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.