The Praeger Handbook of
AMERICAN HIGH SCHOOLS

Volume 1

Edited by

KATHRYN M. BORMAN, SPENCER E. CAHILL,
and
BRIDGET A. COTNER

PRAEGER

Westport, Connecticut
London

ASSESSMENT, CURRICULAR. American high schools have long separated students, often using tests, and then provided them with different educational experiences and opportunities. This entry examines these interrelated practices of **assessment** and curricular differentiation.

Differentiation of **curriculum** occurs when educators make "different knowledge available to different groups of students" (Oakes et al., 1992, p. 570). Curricular differentiation need not entail the stratification of educational opportunities but commonly does, tracking (or ability grouping) being the classic case. Common grouping categories also include English language ability, special needs status, and giftedness.

Curricular differentiation serves the educational purpose of attempting to respond to individual student differences. To do so, such differences must be identified. This is the role played by assessment—the measurement of individual differences. Starting with the premise that individuals arrive at schools with meaningful and classifiable differences in how they should be educated, policy makers have separated and stratified students and opportunities, and using assessments for technical and principled support for those decisions.

However, the function served by assessment and curricular differentiation practices have often gone beyond legitimate educational purposes. They have a disagreeable history marked by racism, classism, and anti-immigrant sentiments. During an era when many policy makers believed that the **intelligence** and societal contributions of ethnic and racial

groups could be hierarchically arranged, testing reinforced those beliefs. Further, the practice of differentiation through sorting and stratification justified the provision of higher-quality educational opportunities to children considered more deserving and more likely to make productive use of their education.

In the earliest years of the United States, the only students who attended secondary schools were those seeking preparation for **higher education**. These schools, generally private, served a fraction of the overall population. Public **high schools** began to develop later, in the years after the Civil War; but by 1890 fewer than 10 percent of 14- to 17-year-olds were enrolled in any secondary school (Oakes, 1985). Such selective enrollment obviated the need for other forms of curricular differentiation.

However, a changing American population and industrialization began to transform American education, starting around 1880. **Immigrants** from Southern and Eastern Europe presented policy makers with a new challenge: Americanization (Tyack, 1974). They asked the public schools to lure the children of immigrants away from the customs, language, and other attributes of their parents. At this same time, Americans from poor, rural areas were migrating to the cities. Social reformers wanted schools to help address the squalor developing in urban areas; businesses were looking to the public schools for a literate work force; and colleges and universities were pushing for a more standardized preparatory curriculum. Schools that had once served only elites were now confronted with students from different backgrounds; their structure and mission changed accordingly: "The solution ultimately settled upon was the comprehensive high school—a new secondary school that promised something for everyone, but, and this was important, that did not promise the *same* thing for everyone" (Oakes, 1985, p. 21).

Industrialization had a two-fold influence during this period. Urban industrialization shaped housing, workforce, and economic opportunities, as well as the role of schools within that context. It also shaped societal values and beliefs. In particular, scientific management styles of industrial efficiency became models for efficiency in schooling, and the scientific management of schools provided fertile ground for the scientific testing of students. Schools felt enormous pressures to match the modern management of industry, and seemingly scientific assessment was a necessary component of doing so.

Scientific testing arose within the larger context of human intelligence research (Lemann, 1999; Gould, 1981; Oakes, 1985). As early as 1869, close to forty years before the first IQ tests, scholars such as Francis Galton argued that intelligence was inherited. Of particular concern was the growing population of "darker skinned" races who were presumedly of inferior intelligence (Lemann, p. 23). Galton, among others, argued for eugenics, or the practice of selective breeding in order to improve intelligence.

In 1905, Alfred Binet developed the first IQ test, a technology that quickly found its acceptance among the intelligence researchers studying the characteristics of different populations and among efficiency reformers looking for scientific ways to identify individual differences in abilities and skills (Gould, 1981; Lemann, 1999). Binet asked children a series of emotionally based questions and then ranked each child's mental age based on her or his answers. The ratio of mental and physical age resulted in an IQ score. During World War I, the U.S. Army administered such IQ tests to over 2 million soldiers, using the test results to make personnel assignments. Population intelligence researchers, in turn, used the results of the Army's testing to determine inferior and superior races.

School reformers saw in this same technology an opportunity to serve all students within the new, comprehensive high schools. Through curriculum tracking, a "common education available to all became an equal opportunity to take differentiated courses to prepare for differentiated roles.... That **inequality** of economic and social position would result was justified by the fairness of students' chances to compete for the most advent aged positions—a fairness undergirded by the science of testing" (Oakes et al., 1992, p. 581). Beginning in the 1920s, and continuing into the 1960s, IQ tests were widely used for these track placements (Shepard, 1992).

For most educators and scholars, the tracking that Oakes and others describe is most troublesome as applied to those students who are enrolled in low-track classes. Researchers have repeatedly found these classes to have lower expectations and fewer resources and to often be little more than warehouses for children that the educational system has cast off.

When is it unfair to enroll students in classes having such a watered-down curriculum? Most Americans perceive a placement as unfair if it is made arbitrarily or if the separation of students appears to be based on unsavory criteria, such as gender, race, or family wealth. In contrast, such enrollment is seen by many as fair if the classes are chosen by the students or their families—an issue we discuss later in this entry. Also, many Americans perceive such enrollment as fair if students' demonstrated achievement and academic capacity merits the weaker curriculum.

This latter rationale is grounded in a straightforward argument. If students enter school with stratified innate capacities and stratified innate potential, then it seems equitable and reasonable for school policies to sort these students into different curricular pathways. Moreover, even if students have stratified capacities and potential that is not innate but is nevertheless unalterable by the school, such sorting may be reasonable. This perspective prompted James Conant, the father of the comprehensive high school, to recommend that high schools make available two different tracks, one track for those students with an IQ score above 115 and a second track for those students scoring below 115 (Lemann, 1999, p. 120). Testing, from this perspective, allows schools to tap into objective merit or capacity. As we shall see, this view is—at least implicitly—still held by policy makers today.

Tests, therefore, provide an ethically useful disconnect between societal inequalities and certain schooling decisions that effectively ration opportunities. They provide an indispensable mediating factor, keeping alive the American myth that each new generation starts afresh with unlimited opportunities—the only limitation being one's abilities. Even though test scores are highly correlated to family wealth and to ethnicity and race, they are generally accepted as objective and fair. The historical transition toward tests is described by Oakes (1985):

> At first, students were openly classified into various programs by their ethnic, racial, and economic backgrounds. This procedure, supported as it was by social Darwinism and notions of the special needs of groups less fit for academic education, was considered scientific, efficient, and egalitarian. But by the end of World War I, this blatantly class-biased assignment of children to different educational programs was being called into question since it so clearly conflicted with the American rhetoric of an open and classless society.
>
> The development of IQ tests lent an air of objectivity to the placement procedures used to separate children for instruction. With the introduction of these tests into schools, "ability" groups came into being. Because the tests were seen as scientific and used sophisticated statistical procedures, they were considered both "objective" and "efficient" means of assigning students. (Oakes, 1985, p. 36)

In short, policies incorporating stratified opportunities need a legitimate basis upon which to claim that they are objectively determining different capacities. Testing provides that basis and does so with results that reflect long-standing racist and xenophobic beliefs concerning which children deserve and which are not.

In the Jim Crow era of overt segregation, policy makers did not bother with the pretense that achievement and ability were the grounds on which children were being sorted. Genetic intellectual inferiority, supported by IQ tests, was an adequate and accepted criterion for the common practice of separating children into different schools. In particular, policy makers created inferior black or Mexican schools designed to teach these students the vocational skills necessary to be successful in careers appropriate for their race and class (Anderson, 1988; Gonzalez, 1990).

By the mid-1960s, policy makers were beginning to move away from such overt racial classifications. So "tests were used by Southern schools resisting desegregation, as a way to resegregate black students into lower tracks" (Heubert and Hauser, 1999, p. 32). As recently as 1997, a federal judge in Chicago sanctioned such stratification as long as the school district, in making course placements, was willing to comply "rigorously with objective criteria, such as scores on achievement tests" (Welner, 2001, p. 86). Since such test scores reflect true merit, any underlying discrimination in society or school is washed away by the calculation of the score. Any legitimate factors that might also be considered are pushed aside by the score. And, most importantly, the scientific determination of the test score creates an aura of equanimity and inevitability that obscures any questions that might be raised about the fairness or educational soundness of a policy that stratifies educational opportunities.

The Aura of Science. Advocates of widespread testing have long benefited from a modernistic belief among Americans in the benefits of scientific progress. Advances in science could solve any newly encountered problem. Moreover, few things add meaning more convincingly than attaching a number to an assertion. "Bob is doing poorly in school" is less convincing than "Bob scored at the 33rd percentile." Accordingly, Americans readily welcomed testing, a science that attaches numbers to powerful social constructions like intelligence and merit.

Test scores also benefit from the belief that they are easily understood, and they have often become shorthand for other concepts. Intelligence, for instance, has become virtually synonymous with IQ test scores. Also, the minimum competency testing movement (from about 1975 to 1985) equated test scores with basic literacy and numeracy. Rankings derived from international test score comparisons, such as the SIMS and TIMSS math and science tests administered in various countries throughout the world, measure the quality of American education against our economic competitors. States' standards-based, high-stakes testing and accountability systems, now operating under the federal **No Child Left Behind** (NCLB) guidelines and sanctions, pronounce the success and worth of schools and, indirectly, teachers.

Because assessment carries the imprimatur of science, the judgments tied to these tests are likewise definitive. Such beliefs underlie and buoy NCLB; if a school's scores are low, then the school must be of low quality. Since this judgment is scientific and definitive, it is reasonable to attach high-stakes consequences. Perceptions of the scientifically based fairness of testing have also been key to promoting policies of grade retention policies, of separate facilities for gifted-identified children, and of school admissions requiring high test

scores. In each case, students are separated into groups based—in whole or in large part—on their test scores and then given different educational experiences and opportunities.

The Increasing Role of Choice. Important schooling decisions are increasingly made by students and their parents. In the past, the location of a student's residence in the local school's catchment area determined the student's public school assignment. In many districts, this has changed, with the growth of open enrollment policies, **magnet schools**, charter schools, and other forms of public school choice.

Through a predictable mechanism, these school choice policies, in conjunction with testing and accountability policies, have resulted in an increase in between-school curricular differentiation. Socio-economic status highly correlates with test scores as well as with the likelihood of actively choosing a school. Also, families with higher socio-economic status tend to choose schools with higher overall test scores, as advertised through "school report cards" published in local newspapers and on the Internet. The result is a self-reinforcing process, with higher-scoring schools enrolling increasing numbers of high-scoring students, and lower-scoring schools losing those higher-scoring students. Curriculum differentiation naturally follows from (and buttresses) this stratification, in a manner analogous to within-school tracking.

School-tracking policies have also been infused with new choice elements. Schools that once decisively assigned students to particular classes now encourage students and their parents to choose their own courses. Many schools do still use test scores, grades, and teacher recommendations to select students for the highest tracks, but the trend is toward choice. Under a typical policy today, the school would tentatively place a student in a given tracked class, such as "Ninth Grade College Prep English." The student or her parents could then opt out of this placement, selecting a more or less challenging course. Empirical studies have shown that these policies with greater choice yield results very similar to older, more rigid tracking policies (Lucas, 1999). The "tracking hierarchy" is reinforced by expectations among educators as well as the students themselves (Yonezawa et al., 2002, p. 38).

Today's tracks remain stratified by race and socio-economic status. Their fairness is now bolstered by three interconnected features: testing, choice, and preexisting expectations. Walking into a low-track classroom disproportionately filled with low-income students of color, an observer might reason, "Well, the students at this school all have the option of choosing their own classes," or "Well, the average test scores of this class are substantially lower than the higher track classes." In addition, although the observer is less likely to articulate the thought, she might also be thinking, "It's natural that a lower-track classroom has a disproportionate number of low-income students of color." Given these three observations, she would have little reason to question the fairness of this tracking.

Gifted and Special Education Classification. IQ tests have survived their historical linkage to erroneous assumptions about racial superiority. Assessments such as the Wechsler Intelligence Scale for Children are regularly used today to identify children as gifted. In some school districts, these tests are the only basis for identification. Other districts may include IQ tests but also assess a child's nonacademic gifts and talent or assess academic abilities, such as spatial reasoning, that are unlikely to be measured in an IQ test.

The Wechsler IQ test is also widely used as part of the **special education** identification and classification process. Such use remains common even though courts and statutes have attempted to curb abuses. In the early 1970s, courts in California considered whether school officials improperly used IQ tests to classify large numbers of African American and

Latino children as "Educable Mentally Retarded" (EMR), placing them in separate EMR classrooms. One court finding that racial differences in test scores for African American children were due to cultural bias in the tests. Another court ruled in favor of parents of a Spanish-speaking student, deciding that children cannot be placed in special education on the basis of tests given in a language other than the child's native language.

Largely as a result of such cases, the law currently provides certain safeguards when testing a child with potential disabilities. The federal **Individuals with Disabilities Education Act** (IDEA) and other federal statutes and court cases have set forth a set of standards: testers must administer evaluations in a nondiscriminatory manner, free from racial or cultural bias; tests must evaluate a specific area of educational need, rather than a single intelligence quotient; evaluators must use more than one procedure in determining whether a child has a disability; and test instruments must be in the child's native language. Although intelligence tests are still used for special education placement, they are now combined with other test instruments that evaluate mental processing and academic skills.

See also Assessment; Assessment in Reading and Writing; Standards-Based Reform.

FURTHER READING

Anderson, J. (1988). *The Education of Blacks in the South, 1860–1935*. Chapel Hill: University of North Carolina Press.

Gonzalez, G. (1990). *Chicano Education in the Era of Segregation*. Philadelphia: Balch Institute Press.

Gould, S. J. (1981). *The Mismeasurement of Man*. New York: W. W. Norton.

Heubert, J. P., and Hauser, R. M. (Eds.) (1999). *High Stakes: Testing for Tracking, Promotion, and Graduation*. National Research Council, Washington, DC: National Academy Press. Available online at http://www.nap.edu/books/0309062802/html/index.html (see pages 31 et seq).

Lemann, N. (1999). *The Big Test: The Secret History of the American Meritocracy*. New York: Farrar, Straus and Giroux.

Lucas, S. (1999). *Tracking Inequality*. New York: Teachers College Press.

Oakes, J. (1985). *Keeping Track*. New Haven: Yale University Press.

Oakes, J., Gamoran, A., and Page, R. (1992). Curriculum differentiation: Opportunities, outcomes, and meanings. In Philip Jackson (Ed.), *Handbook of Research on Curriculum*, pp. 570–608. New York: Macmillan.

Shepard, L. A. (1992). Uses and abuses of testing. In Marvin C. Alkin (Ed.), *Encyclopedia of Educational Research, Sixth Edition*, pp. 1477–1485. New York: Macmillan.

Tyack, D. (1974). *The One Best System: A History of American Urban Education*. Cambridge, MA: Harvard University Press.

Welner, K. G. (2001). *Legal Rights, Local Wrongs: When Community Control Collides with Educational Equity*. Albany: State University of New York Press.

Yonezawa, S., Wells, A. S., and Serna, I. (2002). Choosing tracks: "Freedom of Choice" in detracking schools. *American Educational Research Journal* 39 (1): 37–67.

KEVIN G. WELNER AND KRISTY MARTINEZ