

EXCELLENCE AND EQUITY IN
EDUCATION: HIGH STANDARDS FOR
HIGH-STAKES TESTS

Arthur L. Coleman*

INTRODUCTION

Accountability has become the central issue confronting public education in the 1990s. Echoes of "are my child's teachers qualified?," "does my child's school measure up to others?" and "does my child know enough?" reverberate in conversations about the performance of our schools nationwide. The "national crusade" for high standards in education¹ has focused, in large part, on the need to improve teacher training, provide more resources to schools, and establish clearer and tougher standards for student achievement.² The national dialogue regarding the range of standards-based educational reforms must, however, be equally focused on how to ensure that when we speak of "high standards for all students," *all* means *all*. In this context, one subject of frequent debate is that of high-stakes testing.³ Although few educators dispute the impor-

* Deputy Assistant Secretary for Civil Rights, U.S. Department of Education. This article is based on presentations made at the Equal Education Under the Law Symposium and the National Academy of Sciences colloquium, Excellence and Equity in Education: From Promise to Reality, in February of 1998.

¹ President William J. Clinton, State of the Union Address (Feb. 4, 1997).

² See Goals 2000: Educate America Act, 20 U.S.C. § 5801 (1994); Elementary and Secondary Education Act, 20 U.S.C. § 6301 (1994); National Council on Education Standards and Testing, *Raising Standards for American Education: A Report to Congress*, the Secretary of Education, the National Goals Panel, and the American People (1992); National Governors' Association, *From Rhetoric to Action: State Progress in Restructuring the Education System* (1991).

³ See generally Board on Testing and Assessment, National Research Council, *High Stakes: Testing for Tracking, Promotion and Graduation 2-10 - 2-12* (forthcoming 1999) [hereinafter National Research Council, *High Stakes*]; *infra* notes 27-36 and accompa-

tance of accountability in education, vigorous disagreements surface about the issue of how best to ensure accountability for students, teachers, and schools—especially when the subject of test use arises.

High-stakes tests have been a part of the American education experience for more than a century, and much of the controversy regarding test use is not new. "As in the past, . . . everyone may agree that testing can be a wedge, but some see the wedge forcing open the gates of opportunity while others see it as the doorstop keeping the gates tightly shut."⁴ The context for the exchange about high-stakes tests is different in 1998 than it was as recently as a decade ago. "The reform movement of the 1990s [has endeavored] to move schools to engage all students in learning challenging content and skills in preparation for an adult life where the demand of active citizenship and employment will require people who have both basic and advanced knowledge and skills."⁵ In the last decade, schools increasingly have moved from using high-stakes tests as indicia of minimum or basic competency to use of such tests as benchmarks for high standards learning.⁶ This development has not diminished the presence of high-stakes tests on the

nying text. In this article, the term "high-stakes test" will be used to refer to standardized tests that are administered to measure student learning as a condition of promotion or graduation. There are other kinds of tests that are "high-stakes" because they have significant consequences for students individually, such as achievement and aptitude tests that affect student placements in gifted and talented or special education programs. See, e.g., *Larry P. v. Riles*, 793 F.2d 969, 974 (9th Cir. 1984). Although many of the principles set forth in this article apply to those instruments as well, the focus of this discussion is upon tests used to measure student knowledge and learning, and to make promotional decisions about individual students based on those assessments.

⁴ Office of Technology Assessment, *Testing in American Schools: Asking the Right Questions* 6 (1992).

⁵ Marshall S. Smith & Jessica Levin, *Coherence, Assessment and Challenging Content*, in *Performance-Based Student Assessment: Challenges and Possibilities: 95th Yearbook of the National Society for the Study of Education* 104, 108 (Joan B. Baron & Dennie P. Wolf eds., 1996).

⁶ Board on Testing and Assessment, National Research Council, *Equivalence and Linkage of Educational Tests 2-7* (forthcoming 1999) [hereinafter *Equivalence and Linkage*]; National Research Council, *High Stakes*, *supra* note 3, at 2-7. In the mid-1980s, 33 states had mandated minimum competency exams, 11 as a condition of high school graduation. Office of Technology Assessment, *supra* note 4, at 13.

educational landscape. To the contrary, reforms that are underway increasingly rely on high-stakes tests as a measure of student achievement and school accountability.⁷ There is, accordingly, a need to understand the legal and educational principles that must guide the development and implementation of these high-stakes assessments that are an integral part of the educational reforms in the United States today.

Critical to the success of standards reform and, ultimately, to the success of our nation's youth, is the question of how tests are designed and used as part of standards reform efforts. Indeed, in an era in which educators are appropriately insisting on high standards for students, we should demand no less of the educators who design and implement the standards-based reforms that include high-stakes tests intended to measure student achievement.⁸ We should insist upon high standards for educators, psychometricians, and policy makers who together design, implement and use high-stakes tests.

Determining if, and how, high-stakes tests are to be used in elementary and secondary schools is a complex endeavor.⁹ This is true not only because of the many facets of educational policy implicated in strategic questions that must be addressed in the implementation of high standards reforms, but also because of the unavoidable intersection of educational policy with the law. As educators grapple with ways to accurately measure the performance of their schools, teachers and students, and as the educational and psychometric communities continue to struggle with some very basic issues of test use, the law is never far behind—particularly where individuals may be denied educational opportunities based

⁷ Equivalence and Linkage, *supra* note 6, at 4-8.

⁸ See *Groves v. Alabama State Bd. Of Educ.*, 776 F. Supp. 1518, 1532 (M.D. Ala. 1991) (observing with regard to a challenge to a teacher competency test that "[j]ust as the state expects its teachers to measure up to the more exacting professional demands of today's educational system, it itself must do likewise"). See also National Research Council, *High Stakes*, *supra* note 3, at 2-8 (asserting that establishing and meeting standards for proper test use is of increased importance when the test in question has educational consequences for an individual student).

⁹ See generally *Groves*, 776 F. Supp. at 1519; National Research Council, *High Stakes*, *supra* note 3, at 1-10.

on selected performance measures such as high-stakes tests. A judicial observation in the related area of employment testing explains the challenge in education for administrators, teachers, policymakers, and lawyers alike:

[Testing] is part of the general field of . . . psychology, and possesses its own methodology, its own body of research, its own experts, and its own terminology. The translation of a technical study . . . into a set of legal principles requires a clear awareness of the limits of both testing and law. It would be entirely inappropriate for the law to ignore what has been learned about . . . testing in assessing the validity of [the tests at issue]. At the same time, the science of testing is not as precise as physics or chemistry, nor its conclusions as provable.¹⁰

In education, the intersection of law and policy presents its own unique set of issues, many of which are complex in no small part because of the inherently context- or case-specific nature of the questions presented. There are, nevertheless, certain foundations that should guide any discussion of these issues.

First, the most basic obligation of educators is to meet the needs of students as they find them, with their different backgrounds and abilities, and to inculcate values and teach skills to allow them to grow to maturity with meaningful expectations of a productive life in the workforce and elsewhere.¹¹ Simply put, the job of the educator is to help students to achieve their full potential.¹² This educational mandate is no less present during classroom instruction than it is when educators administer tests and evaluate

¹⁰ *Guardians Ass'n v. Civil Serv. Comm'n*, 630 F.2d 79, 89 (2d Cir. 1980), cert. denied, 452 U.S. 940 (1981).

¹¹ See *Brown v. Board of Educ.*, 347 U.S. 483, 493 (1954) ("[Education] is required in the performance of our most basic public responsibilities, . . . is the very foundation of good citizenship, . . . [and] is the principal instrument . . . in preparing [the child] for later professional training . . .").

¹² Jonathan Fox, *Tests Alone Shouldn't Decide A Student's Fate*, NRC Warns, *Education Daily*, September 4, 1998, at 1 ("Kids should be placed in the kind of educational environment that maximizes their learning.").

and act on students' test results. Premised upon the conclusion that tests should be integral to the learning and achievement of students, one federal circuit distinguished between testing in the employment context and testing in the educational context:

"[I]f tests can predict that a person is going to be a poor employee, the employer can legitimately deny that person a job, but if tests suggest that a young child is probably going to be a poor student, the school cannot on that basis alone deny that child the opportunity to improve and develop the academic skills necessary to success in our society."¹³

Tests, in short, should be instruments used by educators to help students achieve their full potential. Thus, policy makers and the education community must work to guarantee that the establishment of high standards for all students does not unfairly result in the denial of educational opportunity to any one student.

Second, the goals of guaranteeing excellence through the promotion of high academic standards and ensuring that all students have fair opportunities to achieve success in public education are inseparable, mutually dependent goals.¹⁴ U.S. Secretary of Education Richard W. Riley has suggested, in fact, that the "new civil right"¹⁵ in the 1990s is education, and that parents, educators, pol-

¹³ *Larry P. v. Riles*, 793 F.2d 969, 980 (9th Cir. 1984). See also National Research Council, *High Stakes*, *supra* note 3, at 5-6.

¹⁴ See *infra* note 80 and accompanying text.

¹⁵ U.S. Secretary of Education Richard W. Riley, Speech at 1997 National Charter Schools Conference (Nov. 4, 1997). See also President William J. Clinton, Remarks on the 40th Anniversary of the Desegregation of Central High School in Little Rock, Arkansas (September 25, 1997) ("We must not replace the tyranny of segregation with the tragedy of low expectations.").

As used in this paper, the term "civil rights" should be read broadly. In other words, the term is not intended to be limited to civil rights protections that stem solely from anti-discrimination statutes, such as Title VI, which prohibits discrimination on the basis of race, or Title IX, which prohibits discrimination on the basis of sex. See *infra* notes 37, 61-67 and accompanying text. The term is also intended to address the protections that stem from due process guarantees that ensure "fundamental fairness" when new government rules and policies affecting individuals are established. See *infra* notes 41-60 and accompanying text.

icy makers and students should work in concert so that all students receive a quality education based on high standards. Inherent in this view is a recognition that the goals of excellence through high academic standards and fair opportunity for students to achieve success in public education are mutually dependent, inseparable goals—not, as some maintain, competing or irreconcilable objectives.¹⁶

Correspondingly, just as a simplistic “either standards or equity” conceptualization of the issue fails to address meaningfully the educational interests at the heart of the testing debate, so does any discussion that assumes the absence of potential tensions between educational standards and civil rights concerns. The risk exists that schools may maintain lower standards to avoid potential civil rights issues.¹⁷ Indeed, education reform efforts that lead to the establishment of higher standards may well magnify, in the short run, performance gaps among different racial or ethnic

Notably, the anti-discrimination principles discussed throughout this article are generally applicable to disability-related and language-proficiency-related issues. See *Brookhart v. Illinois State Bd. of Educ.*, 697 F.2d 179 (7th Cir. 1983) (addressing discrimination and due process claims of disabled students who challenged the requirement of a minimum competency test as a condition of high school graduation); *Castaneda v. Pickard*, 648 F.2d 989 (5th Cir. Unit A Jun. 1981) (addressing claims of discrimination against students with limited proficiency in English). See also *infra* note 66. See generally Diana Pullin & Perry A. Zirkel, Commentary, Testing the Handicapped: Legislation, Regulations and Litigation, 44 Educ. L. Rep. 1 (March 17, 1988). Given the very distinct conceptual and evidentiary issues relating to accommodations for students with disabilities and for students who are not fully proficient in English, the subject of accommodations will not be covered in detail in this article.

¹⁶ A view among some educators and commentators is that there are irreconcilable conflicts between efforts to promote high standards and accountability in education through the use of high-stakes tests, and efforts to ensure that such tests are fairly used. Ronald Brownstein, in an August 1997 *Los Angeles Times* editorial, succinctly—and erroneously—characterized the standards-equity issue as an “either/or question” and described the standards movement as “approaching a collision” with civil rights law. See Ronald Brownstein, Call for Academic Standards Could Face Test from Civil Rights Law, L.A. Times, August 11, 1997, at A5. Perhaps most frequently overlooked by individuals whose view is that educators may pursue standards or equity (but not both) is the substance of the legal standards that guide federal civil rights laws *and* the educational principles that lie at the heart of policies articulating standards-based reforms. These educational principles are, importantly, part and parcel of the legal inquiry, and they guide legal judgments about high-stakes tests. See *infra* Parts II, III.

¹⁷ Smith & Levin, *supra* note 5, at 121.

groups of students in areas where there are already unacceptable performance disparities among racial groups in the school.¹⁸ Importantly, such reforms may also reflect an existing state of unequal educational opportunity that was previously unknown or otherwise unaddressed. The establishment of high standards, and tests that are intended to ensure achievement at the level of those standards, may provide a basis for educators and policy makers to take a hard look at the best available evidence regarding student performance and to take action to correct inequalities that were previously undetected or ignored.

Instead of pretending that the goals of excellence and equity are inherently irreconcilable, and instead of ignoring the potential tension between these twin goals, the policy makers and the education community should address the issues regarding the use of high-stakes tests as part of standards-based reforms comprehensively and strategically. If educators and policy makers are truly committed to ensuring that all students can learn in an environment in which standards are high, they will work to warrant that the terms "fair and equitable" are synonymous with "high standards." Their challenge is to appropriately address the right questions as the test-related reforms are designed and implemented.¹⁹

To provide a framework for considering ways to meet this challenge, this article will provide in Part I an overview of the state and local educational reforms that are designed to promote high standards and more rigorous measures of student performance in

¹⁸ Jennifer A. O'Day & Marshall S. Smith, *Systemic Reform and Educational Opportunity*, in *Designing Coherent Education Policy: Improving the System* 250, 298 (Susan H. Fuhrman ed., 1993).

¹⁹ The focus of this article is upon efforts to raise standards in the context of elementary and secondary education, and the challenges that those efforts entail. However, many of the legal and psychometric principles discussed in this article are also fully applicable to high-stakes tests used in connection with higher education. For example, the question about the appropriate use of an admissions test to a particular college or university implicates many of the same legal issues and raises a series of questions regarding test use and validity that are explained here. See *infra* Parts II, III. Indeed, one point that transcends the particular focus of this article is this: There is no one-size-fits-all formula that can serve as the basis for determining whether any particular test satisfies legal standards and is educationally appropriate. In both legal and educational terms, the context in which a test is being used, the purpose for which it is being used, and the manner in which it is being administered must be addressed before reaching sound conclusions.

America's public elementary and secondary schools. This article will then examine the issues of fairness and discrimination that arise as educators throughout the nation endeavor to improve student performance and accountability of school systems. Specifically, federal due process and discrimination standards applicable to the use of high-stakes tests that establish benchmarks of student performance will be discussed in Part II. Part III will explore the educational foundations upon which the legal standards are premised and discuss the significant way in which the determination of a test's validity affects ultimate legal conclusions. Part IV will offer some conclusions about the congruence between high standards and equity in the context of current standards-based reforms and the achievability of these goals when standards, curriculum, instruction and tests are aligned. Based on this analysis, educators, policy makers and lawyers can better frame and assess the questions that need to be explored as they consider the use of high-stakes tests in the context of efforts to promote high standards for all students.

I. EDUCATIONAL REFORM: RAISING EDUCATION STANDARDS

The decade of the 1990s has ushered in a renewed national dialogue regarding the need for America's schools to "make the grade." Parents, educators, and other policy makers have joined in the call for better performance by schools, teachers and students. Many have proposed raising education standards to better prepare students for a global economy in which knowledge of technology and the development and understanding of information are essential.²⁰ As our society requires the mastery of ever more complex skills, the challenge that is at the heart of the effort to raise education standards and learning is apparent.

A centerpiece of President Clinton's education agenda is federal support for state and local efforts to improve standards and student performance,²¹ and a proposal to measure student perform-

²⁰ See, e.g., Gerald N. Tirozzi, *It's About Teaching and Learning—Not Testing*, *Educ. Wk.*, August 5, 1998, at 44.

²¹ See *Goals 2000, Educate America Act*, 20 U.S.C. § 5801 (1994); *Elementary and Secondary Education Act*, 20 U.S.C. § 6301 (1994).

ance against a national benchmark of excellence.²² Such federal assistance can provide a clear, nationwide basis upon which state and local education strategies and programs may be assessed and improved, ensuring that education in America will meet world-class education standards in the new century. "The twenty-first century will be a time of remarkable opportunity. With high national education standards, we can make sure all our children have the education they need to seize these opportunities."²³ Most Americans agree with the President and the Secretary of Education in their belief that establishing national standards will help improve public education in America.²⁴

Against the backdrop of apparent consensus that academic standards have been too low and that fundamental change and reform efforts are required to promote higher standards of teaching and learning,²⁵ the national conversation on the issue of high standards has taken many forms in many locales.²⁶ In a society where the fundamentals of education by law and practice are local, there are differences in the ways that states and districts choose to reach the goal of educating all students in an environment of high standards. State assessment programs differ for a variety of reasons. One study recognized as causes for the differences "the educational policy climate in the state, the technical quality issues surrounding the use of assessment to make high-stakes decisions, or

²² See President William J. Clinton, State of the Union Address (Feb. 4, 1997) (proposing a voluntary annual reading test at grade four and a mathematics test at grade eight).

²³ President William J. Clinton, Radio Address (Aug. 30, 1997).

²⁴ See Lowell C. Rose et al., The 29th Annual Phi Delta Kappan/Gallup Poll of the Public's Attitudes Toward the Public Schools, Phi Delta Kappan, Sept. 1997, at 41, 44.

²⁵ Indeed, many state standards remain too low. Even when students are meeting state standards, they are falling behind national and international levels of achievement. See Southern Regional Education Board, *Setting Education Standards High Enough* 2-3 (1997). This study also shows that the state standards of performance are so different from state to state that parents, students and educators cannot meaningfully determine if their children and students have mastered the basics at a level to keep them on a course for college or for success in the employment sector, consistent with educational practices in other states. *Id.* at 4.

²⁶ See generally Linda Bond et al., Council of Chief State School Officers, *Trends in State Student Assessment Programs* 1 (1997).

the status of curricular reform in the state."²⁷ The different strategies share many elements in common, however, especially with regard to testing practices. Twenty-five states in 1996-97 were devising or already implementing "high stakes" assessments to measure progress towards high standards.²⁸ Many school districts and states are requiring that students who otherwise successfully complete their high school curriculum also pass a competency test before they can be awarded a diploma.²⁹ In addition to these graduation requirements, schools are implementing testing requirements for promotion from one grade to the next, as a supplement to the coursework requirements. A county board in North Carolina, for example, has instituted a policy that third- through eighth-grade students who do not attain a designated score on state-administered end-of-grade tests will be retained.³⁰ Many states engaged in these kinds of efforts are also designing and implementing assessment systems that are congruent with their more rigorous academic standards and educational curriculum.

High-stakes tests are not without critics.³¹ Not infrequently, high-stakes tests are used for purposes for which they have not been designed and in ways not contemplated by the test developer.³² Poor testing practices have, in part, led to an array of legal challenges. Education reforms have been challenged based on complaints that school districts have failed to provide the time and

²⁷ *Id.* at 1. See also *Equivalence and Linkage*, *supra* note 6, at tbls. 4-1, 4-2.

²⁸ National Research Council, *High Stakes*, *supra* note 3, at 1-2. See also Council of Chief State School Officers, *Annual Survey of Student Assessment Programs Fall 1997*.

²⁹ See *supra* notes 6-7 and accompanying text.

³⁰ See *Erik V. v. Causby*, 977 F. Supp. 384, 387 (E.D.N.C. 1997).

³¹ See Peter W. Airasian, *State Mandated Testing and Education Reform: Context and Consequences*, 95 *Am. J. Educ.* 393 (1987); Fair Test: The National Center for Fair & Open Testing, *High-Stakes Tests Do Not Improve Student Learning*, *Examiner*, Winter 1997-98, at 1, 4-5. See also Edward Haertel, *Student Achievement Tests as Tools of Educational Policy: Practices and Consequences*, in *Test Policy and Test Performance: Education, Language and Culture* 25 (Bernard R. Gifford ed., 1989) (discussing benefits and negative consequences of testing and making recommendations for responsible test use).

³² See, e.g., *Groves v. Alabama State Bd. of Educ.*, 776 F. Supp. 1518, 1531 (M.D. Ala. 1991). See generally National Research Council, *High Stakes*, *supra* note 3, at 2-1; Office of Technology Assessment, *supra* note 4, at 12-13.

resources to enable students to meet new standards.³³ In a related vein, standards-based reform efforts have been challenged as discriminatory because of the adverse consequences of high-stakes tests on minority students.³⁴ Centered on the high-stakes consequences of tests being used as part of the educational program, these kinds of challenges are premised upon a conviction that such tests will unfairly disadvantage students who have not been provided the necessary educational opportunities to learn the material and master the skills necessary to pass the tests. The claims of discrimination set forth by opponents of the high-stakes tests are buttressed by the huge discrepancy in test scores between African-American or Hispanic students and white students.³⁵ These kinds

³³ See *Bester v. Tuscaloosa City Bd. of Educ.*, 722 F.2d 1514, 1515 (11th Cir. 1984); *Erik V.*, 977 F. Supp. at 386-87; *Crump v. Gilmer Indep. Sch. Dist.*, 797 F. Supp. 552, 553 (E.D. Tex. 1992); *Williams v. Austin Indep. Sch. Dist.*, 796 F. Supp. 251, 252-53 (W.D. Tex. 1992).

³⁴ For example, in a federal court complaint filed in October of 1997, the Mexican American Legal Defense and Educational Fund ("MALDEF") alleges that the use of the Texas Assessment of Academic Skills ("TAAS") test as a condition of high school graduation is discriminatory. MALDEF alleges that white students are almost twice as likely as African-American and Mexican American students to pass the TAAS. See Complaint at 6, *GI Forum et al. v. Texas Educ. Agency*, CA No. SA97CA1278 (October 14, 1997). This complaint has initiated the fourth federal challenge to the TAAS this decade. See *Williams*, 796 F. Supp. at 252; *Crump*, 797 F. Supp. at 553; *In re State of Texas*, OCR Complaint No. 06-96-1021 (1995), discussed *infra* at note 80 and accompanying text. Similarly, in *Erik V.*, the complainant alleged that minority students were more adversely affected by the school district's requirement that students satisfy specific standards as a condition of promotion from one grade to the next. 977 F. Supp. at 389. See generally National Coalition of Advocates for Students, "World Class" Standards Are a Cruel Hoax Without a New Bill of Rights for U.S. School Children, *Educ. Wk.*, Nov. 27, 1992, at 27.

³⁵ Such comparisons lie at the heart of disparate impact cases. See *infra* notes 61-65 and accompanying text. Although the general performance gap appears to have narrowed, in many cases, it still remains. See generally Pascal D. Forgione, Jr., *Achievement in the United States: Progress Since a Nation at Risk?* (Center for Educ. Reform and Empower America, 1998); National Comm'n on Testing and Public Policy, *From Gatekeeper to Gateway: Transforming Testing in America* (1990). See also U.S. Dep't of Educ., *School Poverty and Academic Performance: NAEP Achievement in High Poverty Schools* (1998) (similarly concluding that a "large gap" in academic performance between students in high- and low-poverty schools remains, despite some signs of improvement); Ulric Neisser et al., *Intelligence: Knowns and Unknowns*, 51 *American Psychologist* 77, 97 (Feb. 1996) (discussing differences in performance on intelligence tests among racial and ethnic groups and stressing that despite achievements in the psychometric field "many of the critical questions about intelligence are still unanswered").

of claims add fuel to the fire of those who would argue that the push to high standards further undermines the chances minority students, in particular, have to succeed in their schools and, ultimately, in the workplace.³⁶

Based on descriptions of the legal and educational principles that confirm the interdependence of high standards in education and the principles designed to ensure equity for all students, Parts II and III of this article will establish the foundation for understanding why the position that standards reforms that include high-stakes tests are inevitably harmful to minority students is as untenable as it is flawed.

II. THE LEGAL STANDARDS

Decisions regarding the "fundamental fairness" of testing practices are shaped by due process principles and anti-discrimination protections whose source is the United States Con-

An important segment of discrimination challenges involves the access of students who cannot read, write or understand English to a high standards education, particularly where high-stakes tests measure achievement of those standards. Opponents to such reforms have challenged the fairness of tests that purport to assess the academic skills or knowledge of limited English proficient students who have not had sufficient instruction in English to be able to master a test of knowledge in English. In the MALDEF litigation, *supra* note 34, the complainants also allege that the TAAS discriminates against Mexican American students who comprise a significant portion of the limited English proficient student population in Texas. They allege that unlawful discrimination occurs because the test, given only in English, results in very low passage rates for Limited English Proficient ("LEP") students, "even though many LEP students could exceed the performance levels if the test were given in their home language." Complaint at para. 18.

³⁶ The view that the standards movement jeopardizes educational opportunity for minority students misses the mark in part because of the implication, however unintended, that minority children cannot learn to high standards. Cf. National Research Council, *High Stakes*, *supra* note 3, at 2-3 (asserting that the publication of *The Bell Curve*, that posits that inequality among racial and ethnic groups can be explained by differences in intelligence as measured by tests, is a "recent example" of "[t]he misuse of test data in policy debates"). This view also portends a dual standard of assessment in education and employment so that while it may be acceptable to use tests to make critical life decisions (acceptance to college or entry into the military), it is not important to tell elementary and secondary students, their parents and their teachers how they are doing in mastering the basics early enough so that corrective action may be taken, as needed.

stitution and federal civil rights laws.³⁷ Claims pursuant to these laws have been raised in the context of educational reforms that result in high stakes consequences for students.³⁸ Among the relatively small number of federal opinions that address this issue, most have arisen in the context of minimum standards accountability measures that include high-stakes tests. More recently, reflecting the growing number of standards-based reforms, more federal court opinions and U.S. Department of Education Office for Civil Rights case resolutions have addressed the issue of high stakes consequences for students as part of such reforms.³⁹

Regardless of the precise nature of the claim in these cases, however, the ultimate resolution of these issues has centered upon a determination about the validity of the educational decisions based on results of the high-stakes test in question.⁴⁰

A. Due Process Standards

Federal due process cases that address the use of high-stakes tests center upon the wisdom or legitimacy of test objectives⁴¹; the sufficiency of the notice regarding the imposition of new test re-

³⁷ See U.S. Const. amend. XIV; Title VI of the Civil Rights Act of 1964, 42 U.S.C. § 2000d (1994) (prohibiting discrimination based on race and national origin by recipients of federal funds); Title IX of the Education Amendments of 1972, 20 U.S.C. § 1681 (1994) (prohibiting discrimination on the basis of sex in educational institutions by recipients of federal funds). The implementing regulations for Title VI are located at 34 C.F.R. pt. 100 (1997) and those for Title IX are at 34 C.F.R. pt. 106 (1997).

³⁸ See cases cited *infra* notes 46, 54, 60 and 62.

³⁹ See *id.* See also *In re State of Ohio*, OCR Case No. 15-94-5003; *In re Texas Education Agency*, OCR Case No. 06-96-1021; *In re State of North Carolina*, OCR Case No. 11-98-1070 (pending). These citations to OCR cases, along with others throughout this article, are to decisions or pending investigations by the U.S. Department of Education Office for Civil Rights (OCR). OCR is the federal agency charged with enforcement of civil rights laws that prohibit race, national origin, sex, disability, and age discrimination in educational institutions that receive federal funding. These decisions are not published routinely in a reporting service, but may be obtained upon written request from OCR.

⁴⁰ Validity, as used in this article, is a psychometric term that refers to the accuracy and appropriateness of inferences derived from test results. See generally *infra* notes 72-77 and accompanying text.

⁴¹ See *Erik V. v. Causby*, 977 F. Supp. 384, 389 (E.D.N.C. 1997). This issue is generally termed one of substantive due process. *Id.*

quirements; and, correspondingly, the opportunity students have had to master the materials that will be tested.⁴²

Whether the court addresses a substantive or procedural due process claim, the first inquiry is always whether the student on whose behalf the test use is challenged has been denied a property right or liberty interest cognizable under the United States Constitution.⁴³ Courts generally have held that in states where there are compulsory attendance laws, students have a legitimate entitlement to public education that constitutes a property right to which constitutional protections apply.⁴⁴ Courts have differed, however, in the application of this principle. Although some courts have held that the expectation of a diploma upon high school graduation is a constitutionally protected interest,⁴⁵ the more widely accepted view appears to be that the denial of promotion opportunities or of the opportunity to graduate at a particular time (following completion of required course work) is not a constitutionally protected property interest.⁴⁶

⁴² See, e.g., *Debra P. v. Turlington*, 474 F. Supp. 244, 263 (M.D. Fla. 1979), *aff'd*, 644 F.2d 397 (5th Cir. 1981).

⁴³ See *Debra P.*, 644 F.2d at 404; *Debra P.*, 474 F. Supp. at 266; *Erik V.*, 977 F. Supp. at 390; *Williams v. Austin Indep. Sch. Dist.*, 796 F. Supp. 251, 253-54 (W.D. Tex. 1992); *Bester v. Tuscaloosa City Bd. Of Educ.*, 722 F.2d 1514, 1516 (11th Cir. 1984).

⁴⁴ *Debra P.*, 644 F.2d at 403; *Williams*, 796 F. Supp. at 253.

⁴⁵ *Id.* at 403-04.

⁴⁶ *Bester*, 722 F.2d at 1516 (ruling that plaintiffs failed to establish a foundation for their due process challenge of new promotion standards because "an expectation . . . that the schools would continue to promote students performing in a substandard manner [was] not reasonable and did not form the basis for a property right."); *Erik V.*, 977 F. Supp. at 389 (similar); *Williams*, 796 F. Supp. at 255 (ruling that there was "no . . . constitutional right to receive [a] diploma at a specific graduation ceremony"). See also *Bd. of Educ. of Northport-East Northport Union Free Sch. Dist. v. Ambach*, 457 N.E.2d 775, 775 (N.Y. 1983) (concluding that the expectation of the receipt of a high school diploma was not a cognizable property interest because "functionally handicapped" student plaintiffs were not capable of passing the competency tests at issue and their individual educational plans did not "prepare them to pass the basic skills test"). But see *Crump v. Gilmer Indep. Sch. Dist.*, 797 F. Supp. 552, 554 (E.D. Tex. 1992) (following *Debra P.* and concluding that the state's compulsory attendance law and education program provided the foundation for a constitutionally-protected "expectation that they would receive a diploma if they satisfactorily completed high school").

One federal circuit court has concluded that the right to receive a diploma conferred by state law based on extant academic requirements constitutes a cognizable liberty inter-

In the event that a constitutionally protected interest can be found, federal courts have consistently deferred to educational judgments that the establishment or refinement of high-stakes tests are educationally appropriate—as long as those judgments are deemed reasonable, rational, or not arbitrary.⁴⁷ Improving the quality of education or schools,⁴⁸ ensuring that high school graduates are prepared to compete on a national level and that the high school diploma represents a particular level of achievement,⁴⁹ and encouraging of academic achievement through the establishment of qualitative achievement standards,⁵⁰ have all been recognized as educationally sound bases for the imposition of standards or testing requirements.⁵¹

The federal courts' deference to educational judgment with respect to the appropriateness of having a testing practice is in contrast to the more probing inquiry courts will undertake into the design and use of particular tests. Specifically, courts question whether high-stakes achievement tests are administered appropri-

est under the Fourteenth Amendment where the failure to obtain a diploma may stigmatize a student. See *Brookhart v. Illinois State Bd. of Educ.*, 697 F.2d 179, 185 (7th Cir. 1983). See also *Anderson v. Banks*, 520 F. Supp. 472, 504 n.9 (S.D. Ga. 1981) (suggesting that the publication of denial of a graduation degree may implicate liberty interests). But see *Ambach*, 457 N.E.2d at 776.

⁴⁷ Given the inherent educational judgment that inevitably guides this inquiry, and based on principles of federalism and separation of powers, federal courts, as a general rule, defer to the educational judgments supporting the use of tests to measure student achievement. Federal courts are not school boards and they lack the expertise to substitute their judgment for that of educators. So, although they play a critical role in ensuring that civil rights and constitutional protections are, in practice, fulfilling the promise of equal and fair opportunity, they will not sit to second guess core educational objectives. See, e.g., *Erik V.*, 977 F. Supp. at 389 ("A 'classification' based on students' scores on standardized test[s] is surely the paradigmatic situation for application of rational basis review."). See also *Rankins v. Louisiana State Bd. of Elementary and Secondary Educ.*, 637 So. 2d 548, 555 (La. Ct. App. 1994) (applying rational relationship standard to state's requirement of a graduate exit examination for public schools).

⁴⁸ *Bester*, 722 F.2d at 1515; *Debra P.*, 644 F.2d at 402-03.

⁴⁹ *Anderson*, 520 F. Supp. at 506-07.

⁵⁰ *Erik V.*, 977 F. Supp. at 389.

⁵¹ See also *United States v. LULAC*, 793 F.2d 636, 647 (5th Cir. 1986) (deferring to educational judgments about the value of teacher competency exams, observing "[t]he public interest in teacher competence cannot be gainsaid").

ately and are aligned with the instruction the students have received such that they provide meaningful conclusions about the students.⁵²

Courts address the issue of procedural due process by inquiring about the circumstances in which students have been (or should have been) exposed to the curriculum and instruction on which they are tested.⁵³ The question of notice of the test-related requirements converges with the issue of whether students have had a reasonable opportunity to prepare for the test being administered, based on the instruction received.⁵⁴

⁵² See *infra* notes 53-60 and accompanying text.

⁵³ See, e.g., *Debra P.*, 644 F.2d at 404-06; *Crump*, 797 F. Supp. at 555; *Williams*, 796 F. Supp. at 254.

⁵⁴ See, e.g., *Debra P.*, 644 F.2d at 404-06; *Crump*, 797 F. Supp. at 555. One court has ruled that this protection does not mean that students must have an opportunity to prepare for a specific test that may reflect higher standards than previously imposed. See *Williams*, 796 F. Supp. at 254.

When examining a range of competency tests, courts have struggled with the question of how much notice students should have regarding new academic requirements, sometimes with differing conclusions. Compare *Crump*, 797 F. Supp. at 554 (issuing order requiring high school to allow plaintiffs to participate in graduation ceremony, despite not having passed state proficiency exam which had been administered for seven years, but which was revised with more stringent standards in the academic year preceding the graduation at issue), with *Williams*, 796 F. Supp. at 256 (refusing to issue temporary restraining order requiring school district to allow student to participate in graduation ceremony who failed state proficiency exam which had been administered for seven years, but which was revised with more stringent standards in the academic year preceding the graduation at issue).

Due process violations have been found where there was a 13-month delay between a *de jure* state's imposition of new proficiency test requirements and the denial of a student's diploma, see *Debra P.*, 474 F. Supp. at 264-65, and where there were 12-18 months between the implementation of a new test and the imposition of diploma requirements for disabled students who were not exposed to the "goals and objectives" of the minimal competency test at issue, see *Brookhart*, 697 F.2d at 186. By contrast, no due process violation was found where there was two years between the time of the announcement of the new testing requirement as a condition for graduation and the imposition of the high stakes consequences, and where the school district provided students with an opportunity to retake the test and offered remedial courses. See *Anderson*, 520 F. Supp. at 505. See also *Erik V.*, 977 F. Supp. at 387, 390 (refusing to temporarily enjoin a school grade promotion policy adopted in the summer of 1996 and implemented in the 1996-97 school year that established clear performance standards based on test scores, and that provided for remedial instruction, multiple test administrations, and limited waivers of test requirements). Cf. *Bester*, 722 F.2d at 1515-16 (observing in dicta the

Illustrative of the manner in which courts can be expected to analyze this set of issues is the seminal case to address competency tests in education: the United States Court of Appeals for the Fifth Circuit decision in *Debra P.* In that case, the court ruled that the State of Florida, which proposed to institute a minimal competency test as a condition of high school graduation, was obligated to determine if "the test administered measures what was actually taught in the schools"⁵⁵ The court found that the test lacked content validity in that the State failed to demonstrate that the test in question was a "fair test" of material actually taught in the classroom.⁵⁶ Finding that the test was "fundamentally unfair in that it *may* have covered matters [that were] not taught,"⁵⁷ the federal circuit court upheld the district court's injunction prohibiting the use of the high-stakes test and remanded to the district court for further findings as to whether the schools were actually teaching what was being tested.⁵⁸ In response to this requirement, Florida conducted a four-part validation survey: teacher surveys (sent to 65,000 teachers, of whom 47,000 responded); a survey of all school districts; site visits to follow up on district surveys; and, finally, a small

"troubling" notice of a quality school improvement program that established "specific minimum [reading] levels required for promotion" from grade to grade, where the high-stakes criteria were administered within the same year as the announced improvement goals, but upholding the challenged policy).

As these cases indicate, the inquiry of "how long is enough" to satisfy due process requirements is much more than a simple question of timing. The purpose and use of the test, the context in which it is implemented, and the kind of instructional supports that accompany the test are critical in examining this question.

⁵⁵ *Debra P.*, 644 F.2d at 405.

⁵⁶ *Id.* at 405-06. For the definition of content validity cited by the court, see American Educ. Res. Ass'n et al., *Standards for Educational and Psychological Testing* 9 (1985).

⁵⁷ *Debra P.*, 644 F.2d at 404.

⁵⁸ *Id.* at 408. The failure rate of black students was ten times that of white students, and the state's de jure history was a relevant consideration in the court's conclusion that the use of the test could not withstand legal challenge. The State was unable to justify the adverse racial impact against black students. It failed "to demonstrate either that the disproportionate failure of blacks was not due to the present effects of past intentional discrimination or . . . [that] the diploma sanction was necessary to remedy those effects." *Id.* at 407.

number of student surveys.⁵⁹ In 1984, the federal circuit court lifted the injunction against the test use, concluding that students were sufficiently prepared for the test—based on a determination that the skills tested were included in the curriculum and were recognized by teachers as a subject of necessary instruction.⁶⁰

B. Anti-Discrimination Standards

Claims of race or sex discrimination frequently accompany and are interwoven with the constitutional due process claims that surface in educational testing cases. Generally premised upon federal civil rights statutes and constitutional equal protection principles, these claims are grounded upon the theory that the individual student (or class of students represented) has been denied an educational opportunity or benefit as a result of intentional discrimination in the use of a test, or as a result of race- or sex-neutral policies or practices that have the effect of discriminating against students.⁶¹ In its most general terms, the disparate impact inquiry in-

⁵⁹ *Debra P. v. Turlington*, 564 F. Supp. 177, 180-81 (M.D. Fla. 1983).

⁶⁰ *Debra P. v. Turlington*, 730 F.2d 1405, 1416-17 (11th Cir. 1984). The court did not require direct proof that the material tested was actually taught; instead, the court relied upon indirect and circumstantial evidence to conclude that the test was valid. *Id.* at 1409-13. This conclusion was "bolstered" by the evidence of the State's remedial efforts to assist students in danger of failing the test. *Id.* at 1411.

Although there have been relatively few cases to address the evidence necessary to meet the *Debra P.* standard, federal courts generally have affirmed the point that there must be more than theoretical coverage of the curriculum on such tests. See *Crump v. Gilmer Indep. Sch. Dist.*, 797 F. Supp. 552, 556 (ruling that theoretical coverage of material on a high-stakes test is not the equivalent of evidentiary proof that that material has "actually been taught"). See also *Anderson v. Banks*, 520 F. Supp. 472, 508-09 (construing *Debra P.* to require a "nearly exact match" between the material on a test imposed as a condition for graduation, and the school's curriculum, and finding that a test that "probably contains" items taught in the classroom doesn't satisfy this standard). One court has found that if the test objectives and subcomponents are known to teachers, and if the course curricula is designed specifically to include the material in the test objectives, then there is a likelihood that schools will survive the "fairness" challenge. See *Williams v. Austin Indep. Sch. Dist.*, 796 F. Supp. 251, 254.

⁶¹ The focus of the subject of discrimination in this article is on neutral policies and practices that have adverse consequences on students, based on race, national origin, sex or disability, which are cognizable under federal anti-discrimination statutes and regulations, but not under constitutional provisions. See *Guardians Ass'n. v. Civil Serv. Comm'n.*, 463 U.S. 582, 589-90 (1983).

volves a determination about whether the high-stakes decisions stemming from tests result in an adverse impact on one group (typically identifiable by race, national origin, sex or disability) and not others, and if so, whether there is a sufficient justification for the challenged practice.⁶² In instances in which non-intentional, race- and sex-neutral practices are at issue, a violation of law may occur if (1) a test used to deny a student educational benefits or opportunities has a statistically significant adverse impact upon a group of students based on race, national origin, sex or disability;⁶³

Constitutional claims are cognizable only where there is intentional discrimination, grounded upon evidence of a discriminatory purpose in the establishment of the practice at issue, for which there is no legitimate educational justification. See *Washington v. Davis*, 426 U.S. 229, 248 (1976); *Personnel Adm'r of Mass. v. Feeney*, 442 U.S. 256, 272 (1979). The body of case law dealing with such claims is limited, and the issues regarding high-stakes testing practices in the 1990s do not generally implicate acts of intentional discrimination. See *supra* notes 48-50 and accompanying text. Similarly, while not central to the nature of the high-stakes testing issue present as part of the standards-based reforms, any instance in which the lingering effects of a previously segregated education system exist will lead to additional inquiries and concerns about the imposition of such requirements. The aim of improving educational standards must be considered and analyzed in the context of the school district's *de jure* history, and courts will examine any policy related to student achievement to ensure that it does not perpetuate illegal segregation. See *Anderson*, 520 F. Supp. at 500; *United States v. LULAC*, 793 F.2d 636, 649 (5th Cir. 1986).

⁶² See *Guardians Ass'n*, 463 U.S. at 608 n.1 (holding that Title VI regulations appropriately prohibit actions by recipients of federal funds that produce the effect (or impact) of discrimination, even though the statute itself forbids only acts of intentional discrimination). See also *Alexander v. Choate*, 469 U.S. 287 (1985) (applying the disparate impact analysis in a case of alleged discrimination against disabled recipients of medical care). The use of the impact standard has also been confirmed in the Title IX context. See, e.g., *Sharif v. New York State Educ. Dep't*, 709 F. Supp. 345, 361 (S.D.N.Y. 1989).

Although these actions are modeled after the Title VII employment standard of "disparate impact," see, e.g., *Larry P. v. Riles*, 793 F.2d 969, 982 n.9 (9th Cir. 1984); *Georgia State Conf. of Branches of NAACP v. State of Georgia*, 775 F.2d 1403, 1418 (11th Cir. 1985); *Groves v. Alabama State Bd. of Educ.*, 776 F. Supp. 1518, 1523 (M.D. Ala. 1991), there are very important (and frequently overlooked) distinctions between the employment setting and the educational setting. The questions of validity in the educational context tend to be more complex, particularly given the school's continuing obligation to the student to ensure the appropriate education of the student (including placements based on tests and other assessments). See *supra* notes 11-13 and accompanying text.

⁶³ There is no general or rigid rule that guides the determination of statistical significance in all impact cases. Generally, courts look to percentage disparities, standard deviations, or other statistical formulae to address this component of the impact test. See generally *Connecticut v. Teal*, 457 U.S. 440, 443, 451 (1982); *Groves*, 776 F. Supp. at 1526-29 (comprehensively surveying federal standards and discussing the disparate im-

and (2) the test-related practice leading to the denial of opportunity is not educationally necessary;⁶⁴ or (3) there is a less discriminatory and practicable alternative that as effectively would serve the educational objectives that support the use of the test in the first instance.⁶⁵

pact theory, and the evidence necessary to establish the required statistical disparity, in the context of a challenge to a state board of education's requirement that admissions to undergraduate teacher training programs be conditioned upon minimum test scores).

A frequent misunderstanding relating to these civil rights standards is that the existence of "disparate impact" is itself a violation of the law. The existence of the impact merely triggers the inquiry regarding the educational justification, prompting an examination of the justifications (or lack thereof) of the particular form of and use of the test in question. Any conclusion, for instance, that the federal civil rights laws require the performance of different racial groups to be equal is fallacious. The requirement is that each child have an equal *opportunity* to succeed; it does not require equal results. See *Williams*, 796 F. Supp. at 254 ("Students should be given a fair opportunity to pass the test, not a guarantee that they will pass the test."). A wide range of factors may affect the impact of a test on any group of students, some of which may be divorced from the integrity or validity of the test in question. See *infra* note 79.

⁶⁴ See *Larry P.*, 793 F.2d at 982; *Georgia State Conf. of Branches of NAACP*, 775 F.2d at 1418; *Groves*, 776 F. Supp. at 1530-31. This showing must establish that the challenged practice is "demonstrably necessary to meeting an important educational goal" and that necessity is a substantial legitimate justification for the challenged practice. *Elston v. Talladega County Bd. of Educ.*, 997 F.2d 1394, 1412 (11th Cir. 1993). Accord *Ass'n of Mexican-American Educators v. California*, 937 F. Supp. 1397, 1410 (N.D. Cal. 1996) (requiring employment classifications to have a manifest relationship to the employment question). See also *Larry P.*, 793 F.2d at 982 n.9 (requiring a "manifest relationship" between the test requirement and the educational objective in question); *Sharif*, 709 F. Supp. at 345 (similar).

⁶⁵ The legal determination about the existence of less discriminatory alternatives includes an assessment of cost, timing, and other feasibility factors that relate to the viability of such options. See, e.g., *York v. American Telephone and Telegraph Co.*, 95 F.3d 948, 955 (10th Cir. 1996); *Ass'n of Mexican-American Educators*, 937 F. Supp. at 1426-28 (cost effectiveness); *Sanchez v. City of Santa Ana*, 928 F. Supp. 1494, 1510-1512 (C.D. Cal. 1995) (cost and time expense); *Bridgeport Guardians, Inc. v. City of Bridgeport*, 735 F. Supp. 1126, 1136-37 (D. Conn. 1990) (cost), *aff'd*, 933 F.2d 1140 (2d Cir. 1991); *Sharif*, 709 F. Supp. at 362-64 (difficulty in administration and cost). See also 20 U.S.C. § 6311 (Title I of ESEA) (requiring that state provide for LEP students to be assessed "to the extent practicable, in the language and form most likely to yield accurate and reliable information on what such students know and can do").

In any action where such a claim of discriminatory impact is made, the plaintiff bears the burden of establishing the adverse impact of the high-stakes test; if that *prima facie* case can be established, then the school must establish an educational justification for the test use. Even in the event that a school can establish the necessary justification, a plaintiff may succeed if it can show that there are less discriminatory alternatives that are practicable and that would as effectively meet the educational objectives of the school.

In short, if the test in question is used appropriately, is valid and is the most effective and practicable tool to achieve the school's legitimate purpose (e.g., to promote higher standards in student achievement) then the test will not violate anti-discrimination laws, even if minority students are disproportionately denied educational opportunities because of the use of the test.⁶⁶ "Validity and soundness trump any need to redress discrimi-

York, 95 F.3d at 954; Association of Mexican-American Educators, 937 F. Supp. at 1404; Sanchez, 928 F. Supp. at 1499; Bridgeport Guardians, Inc., 735 F. Supp. at 1130; Sharif, 709 F. Supp. at 361.

⁶⁶ An important complement of claims raising allegations of discrimination issues that presents unique challenges is the use of high-stakes tests for limited English proficient ("LEP") students, students who, by definition, cannot speak or understand English so that they cannot effectively participate in a school's regular education program. Where the inability to speak and understand English excludes students from effective participation in the school's educational programs, a district must take affirmative steps to rectify the language deficiency so that these students will have meaningful access to this instructional program. See 20 U.S.C. § 1703 (1994) (The Equal Educational Opportunities Act) (prohibiting states from denying equal educational opportunities by failing to "take appropriate action to overcome language barriers that impede equal participation by its students in its instructional programs"); *Castaneda v. Pickard*, 648 F.2d 989, 1009 (5th Cir. Unit A Jun. 1981). The reach of Title VI—which prohibits states and school districts receiving financial assistance from the federal government from discriminating against students on the basis of their national origin—extends to limited English proficient students. In 1974, the United States Supreme Court ruled, in fact, that the failure to provide equal educational opportunities to language minority students violated Title VI. *Lau v. Nichols*, 414 U.S. 563 (1974). As with other students, the use of valid test measurements to accurately reflect student achievement is essential. See *Castaneda*, 648 F.2d at 1014.

Anti-discrimination laws do not prohibit the testing of LEP students in content areas in English and making educational decisions about those students based on that information, as long as the underlying educational objective is reasonable and the test in question appropriately furthers that objective. See generally *supra* notes 47-51, 61-66 and accompanying text. Cf. Letter from Richard J. Shavelson, Chair, Board on Testing and Assessment, National Res. Council, to Norma Cantú, Assistant Secretary, Office of Civil Rights, U.S. Dep't of Educ. (June 10, 1996) (on file with the Virginia Journal of Social Policy and the Law).

As with students with disabilities, the provision of appropriate accommodations for LEP students can ensure valid and reliable results. In fact, federal laws require that appropriate accommodations be provided where possible to ensure that conclusions from high-stakes tests administered to LEP students are valid and reliable. See *Castaneda*, 648 F.2d at 1008-09. Accommodations might occur in the test format (including editing accommodations) or in the administration, response or scoring conditions. If students are literate in their native language, and if the instruction has been in that language, providing a valid and reliable version of the test in the student's native language may be an appropriate accommodation.

natory impact."⁶⁷

Federal due process and discrimination opinions related to the use of high-stakes tests point to one question: is the test in question valid for the purposes for which it is being used for all students taking the test? Validity is not a legal term of art. In fact, the legal inquiries, particularly in the discrimination context, focus on the educational justification for the test use in question. They are, consequently, centered upon the validity question precisely because the concept of validation finds its roots in sound educational principles. As Part III shows, the educational underpinnings of these legal inquiries compellingly demonstrate the promise for alignment between the educational objective of attaining high standards and ensuring fairness and equality of opportunity for all students.

III. EDUCATIONAL PRINCIPLES

There are several well-settled educational and psychometric principles that have guided educators and test makers in their efforts to design, administer and evaluate high-stakes tests. Although there may be case-specific disagreements about how these principles should be applied, there is little debate within the educational and psychometric community about their soundness.⁶⁸

First, test results can provide useful, limited guidance in making educational judgments about instruction, placement, and promotional opportunities of students.⁶⁹ Just as tests can provide valuable information about a student's progress in school, they can assist in the comprehensive evaluation of educational programs. Al-

⁶⁷ *Allen v. Alabama State Bd. of Educ.*, 976 F. Supp. 1410, 1431 (M.D. Ala. 1997).

⁶⁸ The overview of educational and psychometric principles set forth in this section is intended to reflect the areas of general agreement relating to high-stakes testing practices. Accordingly, this discussion should be viewed as background for any inquiry regarding the application of federal legal standards to a particular use of a high-stakes test; it is not intended to reflect federal law or policy that would be necessarily applicable in all circumstances. As indicated above, the context in which a test is used and administered guides any determination about the legal sufficiency of the testing practice. See *supra* note 19.

⁶⁹ See Herbert Rudman, *The Future of Testing is Now*, Educational Measurement: Issues and Practice, Fall 1987, at 5-7. See generally Office of Technology Assessment, *supra* note 4; Bond et al., *supra* note 26.

though a test can be a very valuable tool in making decisions about a student's education, a test is just that: one tool among many. And there is no magic to any particular test score. Many factors—some relating to the constructs measured by the test and some not—may affect a student's performance on a particular test. Not surprisingly, therefore, psychometric standards confirm that one test score should not be used as a sole criterion for making high stakes educational decisions about students.⁷⁰ This is particularly true with regard to the interpretation of tests that reflect low scores.⁷¹

Second, the intended use and purpose of any high-stakes test must be established as a predicate to any evaluation of the test design, administration, and use. It is a meaningless exercise to engage in any theoretical assessment of whether a test is valid without knowing the purpose intended.⁷² A test may be valid for one situation or use and invalid for others, just as it may be valid when measuring the performance of some test-takers and not others. And, as new information is assembled regarding a test's use, a conclusion that a test was valid yesterday may yield to information

⁷⁰ See American Educ. Res. Ass'n et al., *supra* note 56, at 54; National Research Council, *High Stakes*, *supra* note 3, at ES-2 (1998) ("[N]o single test score can be considered a definitive measure of a student's knowledge [and] educational decision[s] that will have a major impact on a test taker should not be made solely or automatically on the basis of a single test score."); Board on Testing and Assessment, National Research Council, *The Use of IQ Tests in Special Education Decision Making and Planning* 5 (1996) ("An important maxim of appropriate test use is that no single test score should be used to make decisions about individuals."); National Association for Gifted Children Position Paper, *Using Tests to Identify Gifted Students* (1997); Educational Testing Service, *Sex, Race, Ethnicity, and Performance on the GRE General Test: A Technical Report* 6 (1996). See also *United States v. Fordice*, 505 U.S. 717, 736-37 (1992) (rejecting Mississippi's exclusive reliance on ACT scores in making college admissions decisions where the ACT manual proscribed that admissions should be based on a wide array of factors, including high school grades); *Groves v. Alabama State Bd. of Educ.*, 776 F. Supp. 1518, 1531 (M.D. Ala. 1991) ("[T]he ACT is intrinsically unsuited to be used as an absolute criterion . . . to determine future teaching ability, [especially given its] . . . intended role in college admissions decisions."); 20 U.S.C. § 1412(5)(c) (Individuals with Disabilities Education Act) ("[N]o single procedure shall be the sole criterion for determining an appropriate educational program for a [disabled] child.").

⁷¹ National Research Council, *High Stakes*, *supra* note 3, at 6-8.

⁷² *Id.* at ES-2.

suggesting that it may be invalid tomorrow.⁷³ Therefore, the context in which the test is administered ultimately drives the inquiry about the appropriateness of its use and the conclusions derived from the results of its use.

Third, any test used for high stakes consequences must be validated for its particular use in particular situations.⁷⁴ Tests must accurately measure or assess the qualities, capabilities or characteristics of students taking the test and they must be valid if there are educational consequences for the student. In most basic terms, conclusions derived from test results are considered valid if they accurately reflect the knowledge, ability or other construct the test instrument is intended to measure.⁷⁵ The term validity is generally

⁷³ See Samuel Messick, *Validity*, in *Educational Measurement* 13, 13 (Robert Linn ed., 1993) (asserting that validity is an evolving property and validation is a continuing process). See also Lorrie A. Shepard, *Evaluating Test Validity*, 19 *Rev. Res. Ed.* 405 (1993); William H. Angoff, *Validity: An Evolving Concept*, in *Test Validity* 19 (Howard Wainer and Henry Braun eds., 1988); Committee on Ability Testing, National Research Council, *Ability Testing: Uses, Consequences and Controversies* (Alexandra K. Wigdor & Wendell R. Garner eds., 1982); See also Groves, 776 F. Supp. at 1530 n.26.

⁷⁴ Office of Technology Assessment, *supra* note 4, at 24. Equally clear, however, is the fact that the psychometric evidence necessary to support individual classroom assessments differs from that required where system-wide decisions are made. S.E. Phillips, *Commentary, Legal Issues in Performance Assessment*, *Educ. L. Rep.* 709, 711 (citing Frechtling, *Performance Assessment: Moonstruck or the Real Thing?*, 10(4) *Educ. Measurement: Issues and Practice* 23, 24 (1991)).

⁷⁵ Or, in other words, "[d]oes the test do what it claims to do?" Shepard, *supra* note 73. See also *Debra P. v. Turlington*, 730 F.2d 1405, 1409 (11th Cir. 1984) (similar); *Allen v. Alabama State Bd. of Educ.*, 976 F. Supp. 1410, 1420 (M.D. Ala. 1997) ("Generally, validity is defined as the degree to which a certain inference from a test is appropriate and meaningful."); *Anderson v. Banks*, 520 F. Supp. 472, 489 (S.D. Ga. 1981) ("'Validity' in the testing field indicates whether a test measures what it is supposed to measure."); *Larry P. v. Riles*, 793 F.2d 969, 968 (9th Cir. 1984) (similar); Groves, 776 F. Supp. at 1530 (stating that in the employment context, the inquiry into "whether the test is 'valid' [centers upon the question of] whether an applicant's score on the test yields an appropriate and meaningful inference about the applicant's successful performance of the job").

Evidence of validity need not be perfect, in legal or educational terms, as long as it suggests that the test use in question meets professionally accepted standards. See, e.g., *Association of Mexican-American Educators v. California*, 937 F. Supp. 1397, 1420 (N.D. Cal. 1996). Several sets of professional standards have been developed. See, e.g., *American Educ. Res. Ass'n et al.*, *supra* note 56, at 9 (being revised); Joint Committee on Testing Practices, American Psychological Association, *Code of Fair Testing Practices in Education* (1988); Association for Measurement and Evaluation in Counseling and Development, *Responsibilities of Users of Standardized Tests* (1992); National Council on

understood to refer to the accuracy of conclusions drawn from test results and to actions taken on the basis of those conclusions:⁷⁶ "In essence . . . test validation is an empirical evaluation of test meaning and use. It is both a scientific and a rhetorical process, requiring both evidence and argument. Because the meaning of a test score is a *construction* based on an understanding of the performance underlying the score, as well as the pattern of relationships with other variables, the literature of psychometrics views the fundamental issue as *construct validity*."⁷⁷

Measurement in Education, Code of Professional Responsibilities in Educational Measurement (1995).

⁷⁶ American Educ. Res. Ass'n et. al, supra note 56, at 9 ("[t]he inferences regarding the specific use of a test are validated, not the test itself."). See also Messick, supra note 73, at 13; Lee J. Cronbach, *Essentials of Psychological Testing* 150-151 (5th ed. 1990); Shepard, supra note 73, at 406 (explaining that "validity does not inhere in a test" and that the question of test validity must include an evaluation of the "intended consequences" of the test, and in the particular context of its specific use); Allen, 976 F. Supp. at 1423 (stating that validity equals the degree to which inferences are appropriate and meaningful). Often, validity demonstrations will require careful analysis of data according to existing professional standards. This is a complex and specialized endeavor. See supra note 75 and infra note 77.

⁷⁷ National Research Council, *High Stakes*, supra note 3, at 4-1. Construct validity relates to assessments used to measure a particular characteristic, property, skill, ability, capacity, academic achievement, or behavior. The construct validation of a test usually involves a series of studies, using a variety of research methodologies. For example, a statewide proficiency test designed to measure whether students have learned specific skills or gained specific knowledge in order to determine whether they should receive a diploma would be subject to an assessment of the validity of the constructs of its content and whether the constructs measured by the test are properly aligned to the curriculum and instruction to which students have been exposed.

Along with evidence of a test's validity, evidence of a test's reliability over time and among all students should be considered and must conform to accepted professional standards. Reliability is the degree to which test scores are consistent, dependable, or repeatable. For a test to be considered reliable, there should be evidence that the same students, taking the test multiple times with no change in preparation, receive corresponding scores. No test is perfectly reliable and differing amounts of error or unreliability are tolerated, depending upon the purposes for which the test or procedure is designed to be used. Reliability may be affected by the type of assessment procedure at issue, for example, a standardized test versus a performance-based assessment.

Within the constraints of the defined purposes of a test or procedure, it is expected that the assessment will be fair—that is, valid and reliable for all students taking the assessment, and leading to equitable and just results. See National Research Council, *High Stakes*, supra note 3. There must be adequate evidence that the test is measuring the same academic constructs for all students, and that the results are sufficiently precise for all

In the context of a high-stakes test that is designed to measure what a student has learned, this means that the material on the test must be aligned with the curriculum and instruction of the student being tested. In other words, the material tested on an achievement test to which high stakes consequences are attached should be the same material that the student has been taught, providing him with a fair opportunity to learn the material that is being tested.⁷⁸ If the content of instruction and the teaching methodologies provide the student the opportunity to learn the material being tested, the test is more likely to be valid, reliable and fair.

Fourth, the existence of statistically significant disparities in test scores among subgroup populations is at best a cause for further inquiry and examination about the test use and at worst an indication of the invalidity of the use of the test. Group differences in test scores may reflect a range of causal factors unrelated to the construct of the test or they may reflect a problem with the test. Therefore, educators must endeavor to determine their cause, cognizant of the fact that "[n]ot all inequalities are inequities."⁷⁹

students. A "fair test is one that yields comparably valid scores from person to person, group to group, and setting to setting." *Id.* (citing W.W. Willingham, *A Systemic View of Test Validity*, in *Assessment in Higher Education* (Samuel Messick ed., 1998)).

⁷⁸ *Id.*; O'Day and Smith, *supra* note 18, at 272 ("It is not legitimate to hold students accountable unless they have been given the opportunity to learn the material on the examination."); Shirley M. Malcom, *Equity and Excellence Through Authentic Science Assessment*, in *Science Assessment in the Service of Reform*, 313, 318 (Gerald Kulm and Shirley M. Malcom eds., 1991) ("No assessment can be considered equitable for students if there has been differential opportunity to access the material upon which the assessment is based."). See also *Debra P.*, 730 F.2d at 1409 (a proficiency test must test matters that have actually been taught; it is insufficient if the test covers materials that "may have" been taught); *Crump v. Gilmer Indep. Sch. Dist.*, 797 F. Supp. 552, 555 (E.D. Tex. 1992) (matters on a proficiency exam must cover matters actually taught); *Anderson v. Banks*, 520 F. Supp. 472, 488 (S.D. Ga. 1981) (similar); *Williams v. Austin Indep. Sch. Dist.*, 796 F. Supp. 251, 254 (W.D. Tex. 1992) (course curricula designed to cover written proficiency test objectives passes legal muster).

⁷⁹ National Research Council, *High Stakes*, *supra* note 4, at 4-6, 4-7 (The idea that fairness in testing "requires overall passing rates to be equal across groups is not generally accepted in the professional literature. This is because unequal test outcomes among groups do not in themselves signify test unfairness."); U.S. Dep't of Educ., *School Poverty and Academic Performance: NAEP Achievement in High Poverty Schools 3* (1998) ("The effects of poverty on learning are profound."); *People Who Care v. Rockford Board of Educ.*, 111 F.3d 528, 537 (7th Cir. 1997) ("The social scientific literature on educational achievement identifies a number of other variables besides poverty and dis-

IV. A FOUR-PART EXAMINATION REGARDING HIGH-STAKES TEST USE

This discussion of the complementary legal and educational principles that relate to standards and equity issues in the context of high-stakes testing suggests that the challenges facing educators who are designing and implementing strategies to pursue high standards for all students are not simple or easy. Adherence to principles of high standards and equity are, nonetheless, capable of resolution in a manner that simultaneously supports the dual interests that lie at the foundation of any well developed educational endeavor: accountability and fairness. Indeed, the response to concerns about potentially discriminatory consequences of high-stakes tests should not be one in which the test in question is automatically eliminated. There are many ways in which the goal of understanding the level of a student's performance (objectively measured, in part, by high-stakes tests) can be obtained without subjecting students to discriminatory testing practices.⁸⁰

crimination that explain differences in scholastic achievement, such as the educational attainments of the student's parents and the extent of their involvement in their children's schooling.") (citations omitted). See generally Messick, *supra* note 73, at 85 ("As an instance of unintended side effects, the occurrence of sex or ethnic differences in score distributions might lead to adverse impact if the test were used in selection, which would directly reflect on the apparent functional worth of the selection testing. But whether the adverse impact is attributable to construct-relevant or construct-irrelevant test variance or to criterion-related or criterion-unrelated test variance are salient validity issues in appraising functional worth and justifying test use.").

⁸⁰ See, e.g., *Sharif v. New York State Educ. Dep't*, 709 F. Supp. 345, 361-62 (S.D.N.Y. 1989) (preliminarily enjoining the state's sole reliance on SAT scores when determining merit scholarship recipients, and concluding that the consideration of test scores and grade point averages would provide an appropriate and feasible remedy to the disparate impact against females resulting from the sole use of the SAT scores).

Also illustrative of this point is the OCR resolution in *In re State of Texas*, OCR Case No. 06-96-1021. In that case, OCR resolved a discrimination complaint against the Texas Education Agency regarding the use of the Texas Assessment of Academic Skills test (TAAS) (the passage of which was required to obtain a high school diploma) without eliminating the state's use of the test. The Title VI resolution included four central commitments by the state: (1) school districts would be provided the curriculum and instruction necessary to ensure that all students taking the TAAS had a fair opportunity to pass the test; (2) students would have multiple opportunities (eight during their high school years) to take the test, eliminating any possibility that a one-time failure would deprive a student of any chance at obtaining a high school diploma; (3) the state would continue to develop alternative assessment instruments that would be administered in addition to the TAAS (end of course examinations administered at the end of core courses); and (4) in-

Educators can ensure that they meet the goals of educational excellence and legal soundness through standards-based reforms as long as those reforms that may include high-stakes tests involve appropriate attention and action in four interrelated areas: (1) establishing in clear terms the objectives of high-stakes tests; (2) implementing a methodology for test development and administration that provides a sufficient foundation for ensuring accuracy in educational decision-making; (3) designing the substance of the high-stakes tests so that the achievement that is measured is fully aligned with the school's standards, curriculum and instruction for all students taking the test; and (4) assessing results of test administrations over time to monitor trends and performance. The following sections explain these areas related to test design, implementation and use.

A. Educational Objectives

Before any state or district administers any test, the objectives of the testing exercise should be clear: what are the goals for and uses of the test in question? As an educational matter, the answer to this question will guide all other relevant inquiries about whether the test is educationally appropriate. Without knowing the context in which a test is to be administered and the purpose of a test use, one cannot arrive at any conclusion regarding the appropriateness or usefulness of the test as part of the education of the student taking the test. Judgments about a test's validity are impossible, and the conclusions one may draw from the results are meaningless, absent known objectives. In short, there is no such thing as test validity in a vacuum.⁸¹

tervention and remediation strategies designed to assist students in danger of not passing the TAAS would be implemented.

⁸¹ For example, one cannot conclude that a particular high school proficiency test is, as a general proposition, a valid test. A test of student achievement and learning that is fully aligned with school A's curriculum and instruction and that has been validated according to professional standards for such use could be considered an appropriate tool upon which to make promotional decisions about students taking the test. Conversely, that same test, if used as a placement tool in special education or as a basis for determining college admissions, would neither pass legal nor psychometric scrutiny. Nor, for that matter, would it necessarily be appropriate to use in School B for the same general pur-

Correspondingly, the determination about compliance with federal legal standards related to due process and discrimination protections rests, in the first instance, upon the educational judgment about the purpose and use of the test. Central to this inquiry is the issue of whether high stakes consequences⁸² attach to the test results at all. Is the test designed to monitor a student's or school's performance and to inform educators over time about the prospect of additional resources needed? Or, is the test designed as a gatekeeper, so that failure has particular and real consequences for the student taking the test? In the former case, due process or discrimination claims are unlikely to survive initial legal hurdles. Absent a denial of educational opportunity to a particular student, there is simply no basis upon which to invoke federal constitutional or statutory protections designed to ensure fair and equitable opportunities for students. If, however, the test use in question has high stakes consequences, then the ultimate legal inquiry parallels the ultimate educational question: Do I have sufficient confidence in the test results at issue to allow for informed and consequential decisions to be made about students taking the test? In other words, are the conclusions derived from the test results valid?

This issue can be addressed only with reference to three other central aspects of test design, implementation and use: the methodology of test administration and interpretation; the alignment of standards, curriculum, teaching and test content and, correspondingly, the time period in which educators and students are exposed to clearly articulated standards before encountering high-stakes tests; and the evaluation of test results over time.

B. Methodology of Test Administration and Interpretation

There are a series of protections that may be integrated into the administration of a test and the interpretation of test scores to help eliminate the risk of inappropriately denying (or conferring) educational opportunities to students based on their test scores. Those

pose as in School A—absent a determination that material tested was fairly representative of the material taught. See *supra* notes 74-78 and accompanying text.

⁸² See *supra* note 3.

steps include: (1) establishing compensatory or tutorial supports to ensure that all students have the same basic and fair opportunity to master the material tested;⁸³ (2) providing multiple opportunities for test takers to take the test;⁸⁴ and (3) considering academic factors in addition to the test scores that may affirm or challenge the high stakes conclusions derived from the test scores.⁸⁵

These are all highly contextual practices that inform in very real terms the ultimate fairness of any test instrument. These kinds of protections are premised in part on a recognition that tests are not perfect barometers of learning, and similarly, that conclusions based on those results are not error-free. Given the inevitable presence of the human element in teaching and in learning, there is no guarantee that each student will have received all of the instruction necessary to provide a fair chance for success on high-stakes tests, even where there is alignment among standards, curriculum, instruction and high-stakes tests. Nor is there any foundation upon which one could reasonably conclude that all students can demonstrate what they have learned and what they know equally well on paper and pencil tests. The principles of psychometrics in education recognize these points. Simply, there is no magic to any particular test score. Therefore, there should be ample protections in place to ensure that students are not inappropriately denied meaningful educational opportunities as a result of performance during a single test administration or on a single test.

C. Establishment and Alignment of Test Content

As instruments designed to inform students, parents and educators about success in meeting educational goals, high-stakes tests must be developed and administered as part of the educational effort to improve student performance; they must be integrated into the overall educational package. If, then, tests (as measures of out-

⁸³ See, e.g., *Debra P. v. Turlington*, 730 F.2d 1405, 1411 (11th Cir. 1984); *Erik V. v. Causby*, 977 F. Supp. 384, 387 (E.D.N.C. 1997).

⁸⁴ See, e.g., *Erik V.*, 977 F. Supp. at 387; *Anderson v. Banks*, 520 F. Supp. 472, 489 (S.D. Ga. 1981); *Williams v. Austin Indep. Sch. Dist.*, 796 F. Supp. 251, 252 (W.D. Tex. 1992).

⁸⁵ See, e.g., *Erik V.*, 977 F. Supp. at 387. See also *supra* notes 69-71.

comes) are aligned with the curriculum and instruction (the educational inputs) they will inevitably serve as true measures of the learning that is or is not taking place in the classroom. And as demonstrated earlier in this article, the alignment of the instruction and curriculum of an educational system with the content of high-stakes tests establishes the legally sufficient foundation for meaningful achievement of the goal of educational excellence for all students.⁸⁶

As a result, the issue of notice about the policy changes that will have significant consequences for students is paramount in cases where schools are establishing high-stakes consequences to test results. One federal circuit court has observed: "No one could seriously contend that academic requirements could never be changed during the twelve years a child typically spends in school. It is also obvious that [high-stakes test] requirement[s] could not . . . be[] constitutionally imposed a day prior to graduation. Such late notice could serve no academic purpose."⁸⁷

The issue of the timing between the establishment of educational objectives related to student achievement (and corresponding test requirements) and the actual imposition of high stakes consequences attached to student performance is one that merits serious attention, primarily because of the relationship between the notice provided and the kind of opportunity that students have to master the material or skills on which they can expect to be tested. The question of timing, however, cannot be considered without an understanding of the test purpose and use and administration. Indeed, it would be a mistake to suggest that there are magic rules providing that one year between the announcement of a reform policy and the implementation of high stakes consequences linked to tests administered as part of that policy is insufficient, while three years provides appropriate notice and sufficient time to prepare for the test.⁸⁸ The question of "how long is enough" must be part of a broader inquiry that includes information regarding the educational supports provided to help students succeed.

⁸⁶ See *supra* notes 47-60 and accompanying text.

⁸⁷ Anderson, 520 F. Supp. at 506.

⁸⁸ See *supra* note 54.

D. Results

The perfect is not and cannot be the enemy of the good when evaluating and implementing testing practices. Nonetheless, good, psychometrically-sound testing practices and conclusions must guide the development, administration and use of tests designed to further high standards learning. The evaluation of results relating to test use is necessarily a dynamic one: test results can change over time based on a range of factors that may affect those results, such as teaching practices, methods of test administration, and student limitations and needs. As a result, careful monitoring of test inputs and outcomes over time is critical. Importantly, educators should periodically monitor test results to determine if there are significant disparities among student groups, based on race, national origin, gender or disability. Psychometrically and legally, the presence of such disparities should lead to further investigation so that any potential bias or discrimination in the test use can be eliminated.

CONCLUSION

Although every case will be different, given the complexities inherent in any determination of how a test may properly be used, a fundamental point cannot be lost: All students need schools which expect high performance and offer real and meaningful educational opportunities. Rather than simplistically postulating that high standards are in some way inherently at odds with equity principles in the classroom, we should probe more thoughtfully to ensure that the terms educational excellence and educational equity are synonymous. Similarly, the false choices of eliminating high-stakes tests as student performance measures or embracing them to the exclusion of all other educational criteria that should guide the educational program of students miss the point.

The establishment of clear educational standards as part of ongoing educational reforms can facilitate alignment among curriculum, teaching and testing. This effort can help establish the benchmarks against which success at every level of the educational program is measured so that there is a baseline upon which all educational stakeholders may evaluate the performance of all students and the schools they attend. In particular, they can ensure that: (1)

students and their parents know what the students are expected to learn; (2) teachers and other educators know what needs to be taught; and (3) administrators know the kind of professional development opportunities that should be pursued so that teachers can help their students reach the standards of learning. High standards for academic achievement, when coupled with instruction and support that help students reach those standards, can unite students, parents, teachers, administrators, and community residents around the shared goal of improving student performance.

By defining what students should know and be able to do, standards can, therefore, indicate what assessments must measure in order to show achievement. In return, good assessments can make standards count by giving communities a mechanism by which they can hold schools accountable for results.⁸⁹ Indeed, there should be little doubt that an essential component of any success in obtaining such accountability when high-stakes tests are used will be tests that are well designed, administered and used, ensuring a standard of excellence for *all* students.⁹⁰

⁸⁹ Program and Evaluation Service, U.S. Dep't of Educ., Issue Brief: Why Standards? (1997) (on file with the Virginia Journal of Social Policy and the Law).

⁹⁰ Letter from Richard J. Shavelson, Chair, Board on Testing and Assessment, National Res. Council, to Norma Cantú, Assistant Secretary, Office of Civil Rights, U.S. Dep't of Educ. (June 10, 1996) (on file with the Virginia Journal of Social Policy and the Law) ("Because of the importance of linking test design to specific test uses, validation methods must be designed to provide evidence that test results provide a sound basis for inferences and action. Test validation is often costly, but it is a critical undertaking . . ."). See also *Walls v. Mississippi State Dept. of Public Welfare*, 542 F. Supp. 281, 311 (N.D. Miss. 1982), *aff'd*, 730 F.2d 306 (5th Cir. 1984) ("[A]n invalid test cannot measure 'merit'.").

OCR Draft Testing Guidelines

Print:

Monday, December 13, 1999

Education Dept. Softens Its Tone in Latest Draft of Guide to Using Test Scores

in Admissions

By PATRICK HEALY

Washington

The Education Department has toned down a controversial draft guidebook on the

use of SAT and ACT scores in admissions decisions, after releasing a version in

May that college officials denounced as "anti-test" and an assault on merit-based entrance standards. The department plans to release the new draft of

the guidebook this week, and a copy was leaked to The Chronicle.

Authors of the latest draft -- which is being circulated to higher-education

associations here -- have jettisoned much of the language that worried college

officials in May. The new version spells out more clearly when colleges and

schools might be misusing test scores. And some sections that lawyers had found

threatening have been heavily annotated to cite legal precedents, making the

document appear more matter of fact.

Most of all, the tone is different. The new version says that standardized tests

can be a helpful and useful tool in educational decisions, when used correctly.

Several education experts said the theme of the May version was that reliance on

test scores was more trouble than it was worth.

"Tests are fair when they are valid and reliable for all students who take the

tests," the document states in language that did not appear in May. "Similarly,

the scores must not substantially and systematically underestimate or overestimate the knowledge or skills of members of a particular group.

"

Several higher-education representatives said Friday that the latest version was

more to their liking.

"The document really responds to a lot of the questions and criticisms

OCR Draft Testing Guidelines

we raised before," said Terry W. Hartle, senior vice-president for government and public affairs at the American Council on Education. "They have clearly written it in a form that will work for state and local policy makers." The draft guidelines -- prepared by lawyers in the department's Office for Civil Rights -- are intended as non-binding advice on how institutions can use test scores in non-discriminatory ways. The O.C.R. project proceeded quietly until last May, when the office released a draft that pointed up so many legal issues and test shortcomings that higher-education leaders said it would discourage colleges from using standardized tests and possibly inspire rejected applicants to sue colleges over their test-score criteria. The O.C.R. effort has also drawn attention because it suggests ways to examine and reduce racial disparities in admissions -- a problem that many colleges officials feel should be solved by them, not the federal government. Several minority groups, such as black and Hispanic students, have long scored lower on average on the SAT than white students. Many selective colleges use these scores, to varying degrees, to judge students' abilities, because grades alone cannot tell how rigorous an education students received in high school. The O.C.R.'s May guidelines indicated that colleges that heavily relied on the SAT were discriminating against those minority groups, unless institutions could meet a burden of proof -- including evidence that the test was largely irreplaceable. The O.C.R. appeared, to critics, to be gunning for the SAT, sparking a Congressional hearing on the guide. For example, in the May draft, many readers cited one statement as an implicit threat about reliance on tests, and as a proposal for setting a burden for colleges to show that their policies did not violate federal anti-bias laws. "The use of any educational test which has a significant disparate imp

OCR Draft Testing Guidelines

act on members of any particular race, national origin, or sex is discriminatory, and a violation of Title VI and/or Title IX, respectively, unless it is educationally necessary and there is no practicable alternative form of assessment which meets the educational institution's needs and would have less of a disparate impact," the May version reads. That statement is not in the new version. The latest draft clarifies that disparities in scores or admissions offers do not mean a test is biased or a policy is discriminatory. "Test results indicating that groups of students perform differently should be cause for further inquiry and examination, with a focus upon the relevant educational programs and testing practices at issue," wrote Norma V. Cantú, Assistant Secretary for civil rights, in a preface to the current draft. The proposed guidelines go on to frame questions about disparities, educational necessity, and alternatives, but also make it clear that those suing -- and not colleges -- must meet the burden of proof. Most of the opinions in the latest version are footnoted to refer to studies by the National Research Council Board on Testing and Assessment and to the new Joint Standards for educational testing, prepared by three research groups. By footnoting, the O.C.R. may avoid past criticism that the guide's conclusions are only the views of federal lawyers. But with 161 footnotes in 49 pages, the latest draft is wordier, longer, and more cautious -- even annotating seemingly straightforward words like "fairness." ("Fairness⁵² ... The term fairness has a specific set of meanings, as it is used in test measurement," the guide states, referring readers to Chapter 7 of the new Joint Standards.) Citing studies by outside groups, the new version still warns policymakers that they need to show that they are using tests properly.

OCR Draft Testing Guidelines

"It is important for educators who use the test to request adequate evidence of test quality (including validity evidence), evaluate the evidence, and ensure that the test is used appropriately in a way that is consistent with information provided by the developers or supplemental validation studies." The latest guide continues to suggest pitfalls in test use that may also confuse. Low scores can reflect that the test-takers don't know the subject matter well, the draft states. But if that is the result of limited educational opportunities for those students, the use of scores to make decisions would be "problematic."

"The validity of the test scores would be called into question if the test scores are substantially affected by irrelevant factors that are not related to the academic knowledge and skills that the test is supposed to measure," the new draft states. The new version, like the old, takes a dim view of cut-off scores -- such as colleges requiring a minimum test score to be considered for admission or a scholarship. But this version cites outside experts in warning that decisions should not be based solely on cut-off scores. While the O.C.R. lawyers want to produce a guide that is easy to understand and non-threatening, the latest version expands its section on case law to 15 pages and retains some of the legal feel of the earlier version. Several critics of the May draft said it did not specify when or how potential discrimination occurred; for instance, they asked, how many low-scoring applicants from the same group would have to be rejected to raise discrimination concerns about test use? The new version states that no mathematical formula exists to judge when test-based policies might be discriminatory. Rather, the disparity would need to be "sufficiently substantial," and the burden of proof would be on those suing.

OCR Draft Testing Guidelines

Officials and lawyers for several college associations, civil-rights groups, and testing companies could not be reached for comment Friday or said they would wait until this week to talk about the latest draft. Lawyers who were critical of the May draft and the entire O.C.R. project said Friday they had not seen the new version.

Most of those who did comment declined to be quoted for attribution until then.

They generally agreed that the latest version was much improved from the May draft, but, as one lawyer put it, "the devil is still in the details."

Education Department officials said Friday that they would not comment publicly

on the draft until this week.

The O.C.R. has asked college associations to make comments on the latest draft

by mid-January. After that, another revision is possible, which will then be

available to the public for review. The final guidebook is expected in several months.

Today's news | Information technology | This week's Chronicle | Publishing |

Money

Government & politics | New grant competitions | Opinion & arts | International

| People & events

Information bank | Issues in depth | Jobs | Search the site | Site map

About The Chronicle | How to register | How to subscribe | How to advertise |

Privacy policy | Feedback | Help

Copyright © 1999 by The Chronicle of Higher Education

THE WHITE HOUSE

WASHINGTON

DOMESTIC POLICY COUNCIL

FACSIMILE FOR: Eric Liu

DATE: 12.13.99

TELEPHONE: _____

FAX: 6-2878

FACSIMILE FROM: ANDREW ROTHERHAM

TELEPHONE: (202) 456-5575

FAX: (202) 456-5581

NUMBER OF PAGES INCLUDING COVER: 3

COMMENTS: For as can

U.S. Department of Education, Office for Civil Rights

**The Use of Tests When Making High-Stakes Decisions for Students:
A Resource Guide for Educators and Policymakers**

TALKING POINTS FOR SENIOR OFFICERS

Introduction: Release of the Second Draft

- On Wednesday, December 8, 1999, the U.S. Department of Education's Office for Civil Rights (OCR) will release for review, discussion, and comment the second draft of a forthcoming guide concerning the use of tests when making high-stakes decisions for students. "High-stakes" decisions are those that have significant consequences for students, such as placement, promotion, graduation, and admissions decisions. The draft guide provides an overview of existing test-measurement principles and well-established federal non-discrimination laws related to the use of tests for such high-stakes purposes.
- This guide is being developed in close consultation with the education community. The second draft includes many changes based on prior comments from education stakeholders, and OCR welcomes further comments. In addition, OCR will be holding several meetings with education stakeholders beginning next week to discuss the draft in greater detail.

Purpose of the Guide

- The purpose of the guide is to provide a practical resource for educators and policymakers on how to use so-called high-stakes tests in ways that are educationally sound and legally appropriate, and thereby how to promote the complementary goals of excellence and equity in education.

Background

- Throughout the 1990s, our nation has embraced the goal of promoting high standards for all students. Federal non-discrimination laws fully support this goal.
- Tests, meaning various forms of assessments, can play an important role in promoting high standards for all students. Many states and other actors, therefore, are increasing their use of tests, specifically tests for high-stakes purposes. Federal non-discrimination laws support this use of tests when done in valid, educationally appropriate ways.
- In short, there is substantial alignment between sound educational principles and federal non-discrimination laws when it comes to the use of high-stakes tests.
- However, there is currently a lack of guidance concerning established test-measurement principles and legal standards that should inform the use of tests for high-stakes decisions. This guide is intended to provide that vital information.

Conclusion: Next Steps

- This second draft of the guide is being sent to hundreds of education stakeholders for review. Comments are due by Jan. 18, 2000. After these comments are incorporated, the guide will be submitted to the National Academy of Sciences' Board of Testing and Assessment for its review. Finally, the guide will be released once more for final public comment prior to its publication.



UNITED STATES DEPARTMENT OF EDUCATION
OFFICE OF THE ASSISTANT SECRETARY FOR CIVIL RIGHTS

December 8, 1999

Dear Colleague:

Thank you for your input and interest in our work on *The Use of Tests When Making High-Stakes Decisions for Students: A Resource Guide for Educators and Policymakers*. A revised draft of the Guide, incorporating many of the constructive and helpful suggestions received during the past several months, is enclosed for your further review.

We are continuing our outreach with a series of meetings with stakeholders next week, during which we plan to discuss our work on the Guide, the substance of the current draft, and plans for future work on this important resource. We would be pleased to meet with you to discuss this further.

Overview

Although we have made many changes to the Guide, our objective and foundations remain the same. Our objective is to provide educators and policymakers with a useful, practical tool that will assist in their planning and implementation of policies relating to the use of tests as conditions of conferring educational opportunities to students. The foundations for this Guide include the 1999 Standards for Educational and Psychological Testing published by the American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. These standards provide a framework that can inform good decisions related to the use of tests used for high-stakes purposes. (These standards were made publicly available in November of this year, replacing the 1985 standards cited in the earlier version of this Guide.) Existing federal legal standards and principles—derived from decades-old statutes and regulations, as well as the federal court opinions articulating those legislative and executive pronouncements—also remain as foundations for the Guide. (More extensive citations to those authorities than in the previous draft are provided in footnotes.)

Some of the significant changes from the previous draft include:

- A revised and expanded format that provides a conceptual overview of the relevant measurement and legal principles and that separates the detailed and context-laden discussion of measurement principles (Chapter 1) from the legal principles (Chapter 2);
- A more detailed discussion of measurement and legal principles related to good and nondiscriminatory test use practices, in the context of specific frequently made high-stakes decisions affecting students;

Dear Colleague

Page 2

- An expanded discussion of issues specifically relating to students who are learning English in the elementary and secondary setting and students who are disabled; and
- An expanded glossary of terms and set of related references.

Next steps

We are asking that you provide any further feedback that you may have by January 18, 2000, if you have additional comments relating to this Guide that you would like to share before the draft is submitted to the National Academy of Sciences Board on Testing and Assessment [BOTA]. After BOTA's final review, we anticipate making a draft available to the public for review. We will publish a federal register notice to that effect and will have the revised draft posted on our web site. Once final comments are in, we will issue the Guide in final form.

Do not hesitate to call me if you have questions or additional suggestions. You may also contact David Berkowitz at (202) 205-5557 or <David_Berkowitz@ed.gov>.

Thank you again for your interest in this work.

Very truly yours,

A handwritten signature in dark ink, appearing to read "Arthur L. Coleman", followed by a long horizontal flourish line.

Arthur L. Coleman
Deputy Assistant Secretary



"Coleman, Arthur" <Arthur_Coleman@ed.gov>

10/21/99 03:38:55 PM

Record Type: Record

To: Peter Rundlet/WHO/EOP, Andy Rotherham/OPD/EOP

cc:

Subject: FW: Testing resource guide...strategies for rollout, next steps and schedule

I meant to copy you on this when I sent it on Tuesday. Happy to discuss if you want to talk. Art

> -----Original Message-----

> From: Coleman, Arthur

> Sent: Tuesday, October 19, 1999 12:10 PM

> To: Frost, Susan; Smith, Mike; Liu, Goodwin; Johnson, Judith; Cohen,

> Mike; Jovicich, Catherine; Rairdin, Kae; Fleming, Scott; Frank, David;

> Murphey, Rodger; Winston, Judith; Winnick, Steve; Kole, Adina; McLaughlin,

> Maureen

> Cc: Cantu, Norma V; Thornton, Leslie; Wohl, Alexander

> Subject: Testing resource guide...strategies for rollout, next steps

> and schedule

> Importance: High

>

> Friends,

>

> This is a recap of where we are regarding the revisions to the resource

> guide re testing, and plans for moving forward. (A timetable is located at

> the bottom of this message.) Please feel free to email or call with

> reactions and input. Next week, we'll begin extending invitations for

> some of the early November conversations that are discussed below.

>

> We've received extensive input regarding the resource guide from the 40+

> (education, business, parent, testing, advocacy) groups with whom we've

> conferred. We've almost completed our efforts to incorporate much of the

> feedback in terms of scope, clarity, etc. We plan to circulate within the

> Department a draft of the revised guide next week, allowing one week for

> review and comment. We'll then incorporate comments as appropriate and be

> in a position to circulate among our stakeholders toward the end of

> November, prior to forwarding the draft to the National Academy of

> Sciences Board on Testing and Assessment for its final review.

>

> In the meantime, we plan to schedule a series of meetings with groups

> during the first week in November to provide an overview of the comments

> received, and to discuss the changes that can be expected in the draft.

> The new draft will NOT be circulated during the meetings during the first

> week of November; rather, these efforts will be designed to provide a

> 'sneak preview' -- to gauge reaction, conceptually, and to prepare the

> groups for what they'll soon be seeing. (We've had offers of assistance

> by ACE and K-12 test publishers to facilitate convening of such meetings.)
 > We expect to provide a letter to summarize the presentations to minimize
 > the risk of mischaracterization of the meetings. Through this process we
 > can identify likely trouble spots and be in a position to strategize
 > before distribution of the guide.
 >
 > The NAS Board on Testing and Assessment's Forum on Educational Excellence
 > and Testing Equity (funded by the Dept., with Dept. representation by
 > Judith Johnson and me), with broad representation from the K-12, academic,
 > and testing world, has agreed to meet on November 9 in Washington to serve
 > as a focus group for the guide. We will meet all day, receive feedback
 > from this group, and refine the document to address issues prior to the
 > re-release to educational stakeholders later in the month.
 >
 > I plan to follow up with Susan re group meetings, and Kae re possible
 > meetings with Congressional staffers.
 >
 > Thanks.
 > Art
 >
 >
 > Week of Event
 >
 > October 25 Circulation within ED and WH contacts
 >
 > November 1 Meeting with external groups pre-release of the
 > guide
 > Editing, based on ED comments
 >
 > November 8 NOVEMBER 9: Focus group of NAS Forum members in DC
 >
 > November 15 Editing
 >
 > November 22 and 29 Release of Guide to stakeholders
 >
 > December 15 Guide to BoTA, following revisions based on external
 > feedback
 >
 >

TEST USE AND CIVIL RIGHTS

Why Is Test Use Important?

Tests continue to be used in many ways to measure competence in today's competitive world. From elementary through graduate school, test results often serve as a basis for decision-making that affects our youth. The progress of young schoolchildren increasingly is measured by test scores. Efforts to promote excellence in education are frequently associated with the use of high-stakes tests — tests whose results are used to make placement, promotion and graduation decisions, for instance. To best understand tests, all of us — parents, students, teachers, school administrators and policymakers — should understand the kinds of tests used and why and how they are used.

The issue of nondiscrimination in testing and assessment is properly viewed as consistent with standards-based reforms -- the cornerstone of many of the U.S. Department of Education's initiatives. The U.S. Department of Education is committed to the support of high standards and challenging assessments for all students. Nondiscrimination in testing and assessment is essential to ensuring that equal opportunities for educational excellence are provided regardless of race, national origin, or sex. All students need an educational system which both expects high performance and offers real and meaningful educational opportunities. It is critical that high standards for academic achievement be coupled with the necessary instruction and support that help students reach those standards - as determined by valid and reliable assessments.

The improper use of high-stakes tests can violate civil rights laws that prohibit discrimination against students on the basis of their race, national origin or sex. Any use of a high-stakes test must be considered in the context of the educational interests at issue, constitutional guarantees and civil rights laws.

This pamphlet describes civil rights requirements that apply to high-stakes tests. Specific examples of test use and civil rights requirements are discussed throughout this pamphlet and are also included in the questions and answers section at its end. The constitutional requirements are outlined in one question and answer below. For a more complete discussion of the legal principles related to test use and civil rights, please see OCR's *Nondiscrimination In High-Stakes Testing: A Resource Guide*.

What Tests Have Consequences For An Individual Student?

Most students take tests that are prepared by their teachers and designed to measure knowledge of topics covered in the classroom. For younger students, a weekly spelling test would be an example. For older students, a final exam on American government given at the end of the semester is an example. In addition, there are some tests that are given on a large scale to measure the performance of an entire school or school district. In many such cases, individual student scores are not reported to the school, student or parent. Instead, only group scores are reported for the school or for the school district.

Both classroom tests and broad school assessments are important. The focus of this pamphlet is on another type of test: those that are generally given on a state-wide or district-

wide basis and that are used to make educational decisions that have very important consequences for an individual student. Examples of these decisions are:

- whether or not students will be placed in gifted and talented programs;
- whether or not students will be promoted to the next grade or permitted to graduate; and
- whether or not students will be offered such benefits or opportunities as admissions or scholarships to specific colleges and universities, or to vocational education programs.

In cases like these, tests are used by schools to make major decisions about a student's educational future. Because these tests have important consequences for students, they commonly are called high-stakes tests.

Federal laws prohibit discrimination against students on the basis of race, national origin or sex in testing. This pamphlet outlines the relevant civil rights laws and the legal standards, along with frequently asked questions and answers regarding test use. (In addition, although beyond the general scope of this pamphlet, Federal law also prohibits discrimination on the basis of disability in testing and some of the basic requirements in this area are outlined in the question and answer section.)

Federal Civil Rights Legal Standards Regarding High-Stakes Test Use

Different Treatment

Discrimination against students on the basis of race, national origin or sex can occur in the use of high-stakes tests. One form of discrimination is called different treatment. This occurs when students are treated differently solely because of their race, national origin, or sex in terms of how a test is given or how its results are used, absent an appropriate legal justification (i.e., to remedy past illegal discrimination). One example of different treatment would be if a school district uses test scores to place girls in advanced placement math class only if they achieve higher test scores than boys placed in the same class. Another example would be a school district that puts minority group students in a special education program for mentally retarded students based on their scores on an intelligence test but does not assign white students with the same scores to that program, even though all other placement factors are equivalent.

Disparate Impact

A test may be discriminatory even if it is used in the same manner for all students and even if it is given under policies or practices that are the same for all students: it may result in a disproportionate denial of education benefits or opportunities to a particular group of students. Sometimes test scores result in students of a particular race, national origin or sex being denied — in numbers that are very different from their representation in the general student population — such education benefits or opportunities as promotion, graduation or placement. For example, a test would likely have a

disparate impact if it results in placement of 10 percent of the school's total number of minority group students and two percent of the school's total number of white students in special education classes for mentally retarded students. (Whether or not any particular set of numbers results in the type of disparate impact that triggers concern depends on the outcome of a statistical analysis, a topic discussed in more depth in *OCR's Nondiscrimination In High-Stakes Testing: A Resource Guide*.)

It is important to underscore that such disparate impact, by itself, does not mean that discrimination has taken place. Instead, it is merely a red flag — an indication of possible discrimination — that suggests that additional questions regarding test use should be answered. Before deciding if disparate impact discrimination has taken place, a complete set of questions must be asked and answered in a process involving several steps, outlined below.

Disproportionate Numbers of Students

- ❖ First, using our example, has the test resulted in markedly disproportionate numbers of students of a certain race, national origin or sex being placed in a special education class, as compared to the proportion of students of another race, national origin or sex? If the answer is yes, the next step is to determine the educational necessity of the test.

Validity and Reliability

- ❖ Where the test has a disparate impact, the school district must show that the test is educationally necessary. In determining whether a test is

educationally necessary, it must be shown that the test use is valid and reliable. Professionals in the field of testing use professionally accepted standards to assess the validity and reliability of a test in an educational decision-making context. Inferences from a test are valid if research demonstrates that the test measures what it is designed to measure when used appropriately, if the test is being used by the school in a manner consistent with its designed purpose, and if the test results are relevant to the educational decision in question. For a test to be considered reliable, there should be evidence that the same students, taking the test multiple times with no change in preparation, receive corresponding scores. (Additional information about test validity and reliability is set out in the box accompanying this text.)

If a test has a disparate impact and a school district cannot show that the test is both valid and reliable for its particular use, the test will not be found to be educationally necessary. Where a test with a disparate impact is not shown to be both valid and reliable for its particular use and the school district continues to use it in the same way, the district is in violation of Federal civil rights laws in education. If the test has a disparate impact and a school district can show that the test is both valid and reliable for its particular use, the next step is to determine whether there are any practical alternatives to the test in question that meet the school's educational needs.

More Information on Test Validity and Reliability

Professionals in the field of testing assess the use of a test to make educational decisions according to professionally accepted standards. The following set of inquiries introduces the complex process of determining the appropriateness of a test for use in a particular situation. In general, testing professionals ask the following types of questions.

Central inquiries - The central inquiries are:

- Does research demonstrate that the test measures what it is designed to measure for all students who are taking the test?
- Are the results reliable indicators of what the test is designed to measure?
- Is the test being used by the school in a manner that is consistent with its designed purpose?
- Is this measure relevant to the educational decision in question?

Where the answer to any of these basic inquiries is "no," the results are not appropriate for use in the particular situation. For example, let's use the example of a math achievement test designed for use state-wide in making decisions regarding whether a student is adequately prepared to move to the next grade level in math. A school might use this math achievement test in two ways - one way being a valid test use and one way being an invalid use of the test. First, it might use the test results in making decisions about whether to promote students to the next math grade. This use is consistent with the design of the

test. But what if the school uses the test to place students in a gifted and talented program in language arts? This use is invalid because it is inconsistent with the design of the test; the information from the test is not relevant to placement in a language arts gifted and talented program. In the first instance, test use is valid; in the second, it is invalid.

Where a test is being used as the sole criterion to make a high-stakes decision, the test must be designed for this use and there must be evidence indicating that it is appropriate to use the test as a sole criterion.¹ For example, a test designed to measure general intelligence would not be a good test to use as the sole measure in selecting students for a gifted and talented program. Because information provided by the test publisher indicates that the test was not designed for this purpose, this use of the test is not valid.

Achievement exams - Tests can be used in making decisions about whether students have acquired a certain degree of knowledge and skills. For example, a school district might require that

¹ The Standards for Educational and Psychological Testing, which are generally accepted professional standards that guide testing in schools, state that, "[i]n elementary and secondary education, a decision ... that will have a major impact on a test taker should not automatically be made on the basis of a single test score. Other relevant information for the decision should also be taken into account by the professionals making the decision." See American Psychological Association Standards for Educational and Psychological Testing (1985) at 8.12.

students pass a statewide test created to measure knowledge and skills in mathematics and language arts in order to receive a high school diploma. Schools have the obligation to ensure that all students are provided the opportunity to receive instruction that is fully consistent with and aligned to curriculum goals or standards set by the state or district for all students. Additionally, the state or school district that is testing the students has the obligation to ensure that the assessment is aligned with curriculum goals or standards. States or school districts must also be able to demonstrate that students have had enough time and opportunity to learn the material tested. If the instruction is not consistent with the goals or standards and if students have not had a fair opportunity to learn the material tested, schools have the responsibility to correct these problems.

Professionals in the field of testing use professionally accepted standards to assess the validity and reliability of a test in an educational decision-making context. It is important to note that a test is not necessarily valid or reliable merely because it is widely used by other school systems or because the company that developed it has a sound reputation.

For a more comprehensive discussion about technical considerations regarding tests, readers may consult OCR's Nondiscrimination in High-Stakes Testing: A Resource Guide, which includes a list of references on this topic

Alternatives With Less Disparate Impact

- ❖ Even where a test is valid and reliable, there still may be another test, or another way of measuring student achievement — or of measuring a characteristic such as level of proficiency in English or whatever characteristic the school wants to measure — that would serve the school's purpose as well as the test in question, and that would have a less negative impact on students of a particular race, national origin, or sex. If the use of such tests or measures is a practical alternative to the test in question and meets the school's educational needs, these alternatives must be used.

It is a good educational practice for school administrators to review the results of any testing program. If school administrators find that a test results in a disparate impact based on race, national origin or sex, the best educational practice is to inquire about other testing instruments or measures that would serve the school's purpose, be valid and reliable for that purpose, and that would provide all students with equal access to the school's programs and benefits.

Ways To Improve Test Use

Even if use of a certain test is found to be discriminatory, frequently there are steps a school district can take without eliminating the use of the test. For example, a school district can enhance student learning opportunities to help students master the skills and knowledge measured by the test. Or the district can add to its decision-making process such other evaluation standards as grades, teacher evaluations, portfolios containing student work, or even a second and different test. Lastly, the school can revise the test to make it valid and reliable for the purpose for which it is to be used.

QUESTIONS AND ANSWERS

Q. What information about testing may be useful to parents and students?

A. When parents or students talk to school staff about programs or individual progress, they may want to ask about any tests with high-stakes consequences that will be given during the school year. Parents and students may find it helpful to ask the name of each test; what knowledge, skill or ability each test is supposed to measure (for example, math achievement or general intelligence); and when each test will be given. Parents and students should know the school's overall decision - making process, including how the school judges the test's importance, what factors may be used along with the test, how each of these other factors may be weighted, and what the consequences are for students who fail the test. Parents and students may want to find out what remedial work will be offered to the student who performs poorly on the test and determine what additional opportunities, such as optional after-school classes to improve test performance, will be offered. A group of parents or an organization like the P.T.A. may want to meet with school counselors and administrators to learn about school tests. Many school districts find that the best educational practice is for the district to take the initiative in providing information about testing to parents and students as early as possible.

When high-stakes decisions are made, including decisions based, at least in part, on tests, parents and students may need information that will enable them to understand how the decisions were made and to assess whether a student or group of students is being treated fairly, regardless of race, sex, or national origin. In terms of facilitating understanding on the part of parents and students, the best educational practice is for school administrators to be able to explain how the overall decision-making process worked.

Q. What are proficiency and achievement tests?

A. Proficiency tests evaluate the mastery of knowledge and skills. They can include such tests as those which evaluate students' reading and writing skills in English. They might also include tests which evaluate academic mastery in subject areas taught in school, such as mathematics or science. These tests are often called achievement tests.

In connection with the national trend toward increasing accountability and encouraging high standards, many states and school districts are using the results of achievement tests to help determine graduation or grade promotion. There are many reasons for this, including ensuring that high school graduates are prepared to either enter college or compete in the job market. States and districts may also want to motivate students to work toward greater academic achievement, or to ensure that high school diplomas meaningfully represent a particular level of achievement.

Q. Are there any Federal requirements affecting public elementary and secondary school students with disabilities in connection with the use of high-stakes tests?

A. As two of the Federal civil rights laws listed on the inside pamphlet cover note, public elementary and secondary schools are required to provide a free appropriate public education to all students with disabilities in their jurisdiction. To comply with the requirements of these two Federal civil rights laws, a school must provide regular or special education and related aids and services necessary to meet the student's educational needs so that the student can participate in and benefit from the school's education program -- including participating in the public school's testing program.

When students with disabilities are tested, the civil rights requirements discussed in this pamphlet apply to them. In addition, schools must, among other requirements, select and administer the tests so that the results accurately reflect what the student knows or is able to do, rather than the student's disability. This means that students must be given appropriate accommodations and adaptations in the administration of the tests. Examples include oral testing, large print tests, Braille versions of tests, individual testing and separate group testing.

One high-stakes decision that affects some students with disabilities and that may involve tests, as well as other types of information, is the decision as to whether a student should be provided with special education. This decision

involves other decisions including: whether the student is an individual with a disability, covered by Section 504 and Title II; whether the student should be provided regular education with related aids and services or special education; and whether the student would be eligible under the Individuals with Disabilities Education Act (IDEA) (discussed below). Under both the civil rights laws and the IDEA, any determination of whether a student should be provided special education must be made on an individual basis in accordance with specific statutory and regulatory requirements, including requirements regarding the use of tests for that purpose.

The IDEA provides funds to states, and through them to local school districts, to assist in providing a free appropriate public education to students residing within the state in mandatory age ranges, and it establishes conditions for receipt of such funds. Under IDEA, the determination of whether a student needs special education must be made on an individual basis through a process which involves the use of tests or other evaluation materials and procedures. Under IDEA, states also must have nondiscriminatory procedures for purposes of evaluation and placement of students in special education, as well as procedures regarding the participation of students with disabilities in general state and districtwide assessment programs (discussed in the next question and answer). State and districtwide assessments of student achievement cannot be used alone for determining whether a student has an impairment and needs special education and related services under the IDEA. However, an individual student's performance on such an assessment could be considered by parents and school districts as a part of an evaluation for the

purposes of determining the student's need for special education under IDEA. For more information on how IDEA requirements affect high-stakes testing, please telephone the Department's Office of Special Education Programs at 202-205-5507.

Q. Should public school students with disabilities be included in proficiency tests with high-stakes consequences that are given throughout a district or state?

A. Under IDEA, this decision must be made on an individual basis by the student's individualized education program (IEP) team, and must be reflected in the student's IEP. For students who are not covered by the IDEA, but who are covered by Section 504, this decision must be made on an individual basis through other applicable evaluation and placement processes. It would be a violation of the civil rights laws prohibiting discrimination on the basis of disability, if a student with a disability who, based upon his or her IEP or Section 504 plan, should be preparing for and taking a state- or district-wide proficiency test, is excluded from these opportunities on the basis of disability. It is generally expected that students with disabilities should be included in these assessments. As described immediately above, where necessary, appropriate accommodations and adaptations in the administration of the test must be provided to students with disabilities who take these tests and should be specified in the student's IEP or Section 504 plan.

The newly enacted Individuals with Disabilities Education Act Amendments of 1997 (IDEA '97) requires States, as a condition of receiving IDEA funds, to include students with disabilities in State and districtwide assessment programs, with appropriate accommodations, where necessary. IDEA '97 also requires that the student's IEP specify any individual modifications in the administration of State or districtwide assessments of student achievement that are needed in order for the student to participate in such assessment. Similarly, if the IEP team determines that the student will not participate in a particular State or districtwide assessment of student achievement (or part of such an assessment), the student's IEP must include statements of why that assessment is not appropriate for the student and how the student will be assessed. IDEA '97 also requires state or local educational agencies to develop guidelines for disabled students who cannot take part in state and districtwide tests to participate in alternate assessments. These alternate assessments must be developed and conducted not later than July 1, 2000. For more information about the IDEA, please call the Department's Office of Special Education Programs at 202-205-5507.

Q. How are limited English proficiency (LEP) students affected by discrimination laws in the use of high-stakes tests?

A. LEP students ordinarily must be included in school or district assessment programs. When LEP students are included in these assessment programs, the inferences and interpretations drawn from the students' responses to the test or assessment procedure must be valid and reliable. Further, the students must be provided appropriate

accommodations in order to ensure valid and reliable results. Accommodations might occur in the test format (including editing accommodations) and/or in the administration, response or scoring conditions. Depending upon the nature and purpose of the test and the particular needs of a LEP student, providing a valid and reliable version of the test in the student's native language might be an appropriate accommodation. Other accommodations may include extended time or the use of bilingual dictionaries. Title VI requires the inclusion of LEP students in assessment programs, absent an educational or psychometric justification for their exclusion. If students are excluded from assessment programs, comparable information about their academic progress must be collected for these students.

Q. In addition to the civil right laws discussed in this pamphlet, are there any other Federal rights or requirements related to the use of proficiency tests for high-stakes decisions about which parents, students and school staff should be aware?

A. Yes, there are rights and requirements that arise from the Constitution of the United States. These apply to students and their public schools and they somewhat overlap with statutory civil rights requirements. An overview of these rights follows. However, it is important to point out that OCR does not enforce constitutional rights, unless there are claims of discrimination which are covered by the

federal civil rights statutes. For this reason, private legal counsel should be consulted for more information or to seek enforcement of these rights in Federal Court.

The Constitution requires that fundamental fairness be present in situations in which a government institution — such as a public school — creates an understanding on the part of students that they are entitled to something important: for example, a high school diploma. What if a student enters high school and the rules provide only that successful completion of coursework is necessary to receive a diploma? And then, when the student enters the senior year, the rules change to require that students also pass a proficiency test to earn a diploma? In this case, there are key issues related to whether a high-stakes test complies with constitutional standards for due process. Among the key issues addressed by Federal courts in making this determination are:

- whether there is a reasonable educational justification for the test;
- whether the test measures what it is designed to measure;
- whether the test represents a fundamental change in the rules of the school related to an important expectation, such as high school graduation, and, if so, whether the students had an adequate opportunity to prepare for, take, and pass the test.

Q. When should a parent or student file a complaint with OCR regarding test use?

A. School districts may take voluntary action to correct discrimination when it is brought to the attention of school decision-makers. However, parents or students may file a complaint with OCR immediately if they believe that discrimination has occurred.

Q. What are some of the steps OCR takes when investigating a discrimination complaint about the disparate impact of a test used to make a high-stakes decision?

A. OCR seeks information regarding whether the use of the test in question has resulted in a markedly disproportionate number of students of a certain race, national origin or sex being placed in or denied access to a particular program. Next, OCR determines if the school has shown that the test is educationally necessary. In determining educational necessity, OCR examines evidence of the test's validity and reliability, as described above. OCR then determines if there are practical alternatives to the test. Specifically, are there other tests or assessment procedures that would have a less negative effect on the basis of race, national origin or sex; or is there another reasonable way to achieve the school's objectives that would have a less negative impact on groups of students, while accomplishing the same educational objective as the test in question? Where appropriate, OCR also would be interested in the ways in which the school or school district would enhance learning opportunities so that students are prepared for the high-stakes test in question.

Q. How does a student or parent file a complaint with OCR?

A. If a student or a parent, or another person, decides to file a complaint with OCR, the complaint should be filed with the OCR enforcement office responsible for the state in which the school is located. The offices are listed on the last page of this pamphlet. Generally, the complaint should be filed within 180 days of the last act of alleged discrimination.

The complainant should give OCR his or her name, address, and daytime phone number, and provide the date(s) and enough information about the alleged discriminatory act(s) so that OCR can understand the nature of the complaint.

OCR may extend the time for filing a complaint in certain circumstances. For example, when a student has filed a grievance under school procedures within 180 days of the last act of alleged discrimination, OCR will generally accept a complaint raising the same allegations up to 60 days after the end of the school procedures because it encourages students to file grievances with their own school first. How OCR resolves the complaint usually then would be limited to the allegations raised in the grievance.

[END]

Bradley, Bridget

From: Coleman, Arthur
Sent: Friday, June 11, 1999 10:12 AM
To: 'Peter_Rundlet@oa.eop.gov'; 'Martin_T@a1.eop.gov'; 'kmitchell@nas.edu'; 'mfeuer@nas.edu'; Bowers, Susan; Lewis, Cathy H; Lim, Jeanette; Slayton, Lester; Cantu, Norma V; Pierce, Raymond; Bradley, Bridget; Jovicich, Catherine; Wohl, Alexander
Subject: FW: NYT response to Thernstrom op ed going today...

FYI, we're submitting this to NYT today. Call if questions. (Bridget, Mike read a longer version yesterday and liked it; we had to cut to get to NYT designated length.) Art

-----Original Message-----

From: Coleman, Arthur
Sent: Friday, June 11, 1999 10:02 AM
To: Green, Julie; Frank, David
Cc: Murphey, Rodger; Lyon, Tom; Saunders, Kelly; Berkowitz, David; Fitch, Rebecca; Winnick, Steve; Craig, Susan; Kole, Adina
Subject: NYT response

Julie and David,

Here is the revised letter cut to about 290 words. Julie, since we were well under 300, I left in two segments that you can still choose to cut if you think advisable...[1] parenthetical in second line of third para ("and conclusion...") and [2] last line of third para ("The guarantee...)--both of which I think are impt.

Hate losing the SAT para, but no way to pare it down much shorter than it was.

It's yours...! Pls send a copy of whatever the final is so that we can provide to staff. Thanks for your help. Art



NYTThernstromltr3.doc

In her commentary, "Testing, the Easy Target," Abigail Thernstrom makes a number of erroneous conclusions regarding the U.S. Department of Education Office for Civil Rights (OCR) draft resource guide regarding test use.

Our effort is to help policymakers and educators understand the foundations for legal and educationally sound testing policies-to avoid controversy and litigation. With citations to over sixty federal court opinions and test measurement standards of the American Psychological Association, this is nothing new.

Ms. Thernstrom's equation of racial gaps in test performance with discrimination (and conclusion that tests are, therefore, "guilty until proved innocent") reflects a misreading of the draft, just as it reflects a gross misunderstanding of the settled legal principles upon which it is based. Differences in students' test scores may result from a range of factors, including, lack of preparation for or motivation to do well; poor skills or knowledge related to what is being measured; or problems with the use of the test itself. The guarantee under federal law is for equal opportunity-not equal results.

Thernstrom also accuses the U.S. Department of Education's Office for Civil Rights of seeking to ban all standardized testing. That outrageous assertion is squarely refuted by the draft, as well as by our numerous policy statements and resolutions of discrimination claims. Our guiding principle is that anti-discrimination standards give substantial deference to sound educational judgments. As a result, tests that are used in educationally appropriate ways and that are valid for the purposes used are very important instruments to help educators do their job. Our resolutions of discrimination claims recognize clearly that the solution to many such problems is *not* to eliminate the very tools that help provide a meaningful picture of the educational opportunities provided to students.

Arthur L. Coleman
Deputy Assistant Secretary for Civil Rights
U.S. Department of Education

In her commentary, "Testing, the Easy Target," Abigail Thernstrom makes a number of erroneous conclusions regarding the U.S. Department of Education Office for Civil Rights (OCR) draft resource guide regarding test use.

Our effort is to help policymakers and educators understand the foundations for legal and educationally sound testing policies—to avoid controversy and litigation. With citations to over sixty federal court opinions and test measurement standards of the American Psychological Association, this is nothing new.

Ms. Thernstrom's equation of racial gaps in test performance with discrimination (and conclusion that tests are, therefore, "guilty until proved innocent") reflects a misreading of the draft, just as it reflects a gross misunderstanding of the settled legal principles upon which it is based. Differences in students' test scores may result from a range of factors, including, lack of preparation for or motivation to do well; poor skills or knowledge related to what is being measured; or problems with the use of the test itself. The guarantee under federal law is for equal opportunity—not equal results.

Thernstrom also accuses the U.S. Department of Education's Office for Civil Rights of seeking to ban all standardized testing. That outrageous assertion is squarely refuted by the draft, as well as by our numerous policy statements and resolutions of discrimination claims. Our guiding principle is that anti-discrimination standards give substantial deference to sound educational judgments. As a result, tests that are used in educationally appropriate ways and that are valid for the purposes used are very important instruments to help educators do their job. Our resolutions of discrimination claims recognize clearly that the solution to many such problems is *not* to eliminate the very tools that help provide a meaningful picture of the educational opportunities provided to students.

Arthur L. Coleman
Deputy Assistant Secretary for Civil Rights
U.S. Department of Education



UNITED STATES DEPARTMENT OF EDUCATION
OFFICE OF THE ASSISTANT SECRETARY FOR CIVIL RIGHTS

June 21, 1999

Dear Colleague:

I am writing to thank you for your interest and input regarding our draft of *Non-discrimination in High-Stakes Testing: A Resource Guide*, and to confirm some of the foundations and objectives for this work. In addition, I want to let you know of our plans for further work on the Guide. We appreciate your willingness to assist in our effort to refine the existing draft of the Guide.

Overview

As you know, our objective is to provide educators and policymakers with a useful, practical tool that will assist in their planning and implementation of policies relating to the use of tests as conditions of conferring educational opportunities to students. As Secretary Riley said in his commemoration of the 45th anniversary of the *Brown v. Board of Education* decision, the courtroom is not the optimal place to make education policy. If, therefore, we can provide meaningful tools to our constituents in an area that is too frequently the subject of confusion—and, therefore, sometimes unnecessary controversy—we will have helped provide a stronger foundation for better educational decisions that serve all of our students. That is our goal.

Background

The U.S. Department of Education Office for Civil Rights enforces laws that prohibit discrimination on the basis of race, national origin, sex, disability and age by educational institutions that receive federal funds. Our work on the Guide is just one part of our effort to prevent problems of discrimination in the area of test use. Our educational stakeholders at all levels have come to us requesting advice and technical assistance in a variety of test use contexts, particularly as states and districts use tests as part of their standards based reforms. And, increasingly, we are addressing testing issues in a broader and more extensive array of complaints of discrimination filed with our office. These corresponding developments confirm the need to provide a useful resource that will capture legal, test measurement, and educational principles, references, and resources to assist educators and policymakers.

In our effort to develop such a resource, we have consulted with dozens of teacher, administrator, policymaker, business, advocacy, and test publisher organizations to solicit input and advice regarding the scope, framing and kinds of resources to include in the guide. Notably, we have contracted with the National Academy of Sciences Board on Testing and Assessment, which has independently reviewed and which will again assess this document to ensure that it comports with professional standards.

400 MARYLAND AVE., S.W. WASHINGTON, D.C. 20202

Our Mission is to Ensure Equal Access to Education and to Promote Educational Excellence throughout the Nation.

Dear Colleague
June 21, 1999
Page 2

Central principles

The draft of the Guide reflects and we believe that tests can be critical components of educational strategies designed to promote educational excellence for all students. Our aim is to ensure that tests are used appropriately. Indeed, our resolutions of claims of discrimination recognize most clearly that the solution to concerns regarding discrimination is *not* to eliminate the very tools that help provide a meaningful picture of the educational opportunities provided to students. In short, this Guide is not about eliminating tests; rather, the Guide should help promote decisions ensuring that tests are used in ways that are consistent with their design and purpose, and that all students are afforded the opportunity to achieve to high standards.

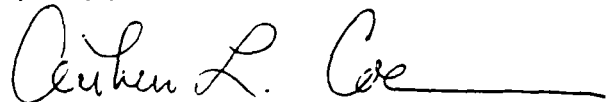
The Guide does not articulate new legal principles or test measurement standards. Citing thirty federal court decisions and over thirty test measurement standards, it provides a synthesis of often very dense information in one place, so that readers who are not as familiar with the area can have a readable, understandable resource.

Next steps

We have already received many useful and constructive suggestions regarding the Guide and are incorporating many of those ideas in our revisions. We have asked all with whom we have consulted to provide us with comments by the end of this month so that we may work during July and August to further refine and develop the Guide. Given the interest in this work, we are pleased to extend this deadline until July 16 for those who need additional time to comment. Our plan is to re-circulate the document in draft to the groups and individuals with whom we have consulted, just prior to our submission of the document to the National Academy of Sciences Board on Testing and Assessment, for its final review. Thereafter, we anticipate making a draft available to the public for review. We will publish a federal register notice to that effect and will have the revised draft posted on our web site. Once final comments are in, we will issue the Guide in final form.

Thank you again for your interest in this Guide, and do not hesitate to call me if you have questions or additional suggestions. You may also contact David Berkowitz at (202) 205-5557 or David_Berkowitz@ed.gov.

Very truly yours,

A handwritten signature in cursive script, reading "Arthur L. Coleman".

Arthur L. Coleman
Deputy Assistant Secretary

United States Commission on Civil Rights
Public Briefing
Prepared Statement of Arthur L. Coleman
Deputy Assistant Secretary
U.S. Department of Education Office for Civil Rights
June 18, 1999

Introduction

Thank you for the opportunity to represent the U.S. Department of Education and to discuss the existing legal, educational and test measurement principles that guide the work of the Department's Office for Civil Rights ["OCR"]. All individuals making important high-stakes decisions affecting the lives of students should understand the central principles on which so many in the education community agree. These points of agreement provide a very fertile common ground that should be the basis of our efforts to ensure that tests are used appropriately and that, as a consequence, accurate educational decisions are made—permitting all students achieve to their full potential.

Our goal is to provide some critical foundations for fulfilling the promise of the new civil right identified by U.S. Secretary of Education Richard Riley in his commemoration of the 45th anniversary of the *Brown v. Board of Education* decision. He said: "A quality education must be considered a key civil right for the 21st century." That point of consensus, along with the common ground that exists regarding good testing policies and practices, provides an important context for this discussion and affirms the need to move from the polarizing, either-or rhetoric that too frequently surrounds this issue. Secretary Riley has noted that too often in education, people are "choosing sides, not solutions." We seek to promote educationally sound solutions through our work related to the use of tests as foundations for high-stakes decisions affecting students.

When we talk about promoting the goal of achieving high standards education for all students, we mean it. The issue of testing in education should not be about favoring either standards or equity. Neither should the issue of testing in education be about blanketly favoring or opposing tests.

We believe that high standards for all means high standards for all. We believe that good test use practices advance high standards learning and equal opportunity—just as educationally inappropriate uses of tests do not. Tests are, as Secretary Riley has said, “important tools for educators to assess and assist students as they strive to meet high standards.” And, as they often provide a meaningful picture of educational opportunities provided to students, our goal is to preserve these critical measures of student performance just as we work to ensure that they are used appropriately.

We believe that the use of tests in education is an issue that should be the subject of informed and constructive dialogue. We welcome the opportunity today to continue our effort to advance a constructive discourse that can result in educational excellence for all students.

OCR's Work Regarding High-Stakes Decisions and the Use of Tests

The recently published *High Stakes: Testing for Tracking, Promotion, and Graduation* (National Research Council, Heubert and Hauser, eds., 1999) observed that the controversy surrounding the use of tests for high stakes decisions affecting students is often based upon misinformation and misperceptions about what tests are designed to do and, correspondingly, about good (and bad) test use practices. The application of federal non-discrimination laws to testing practices is subject to the same fate of misinformation—and sometimes, ill-informed commentary. To promote a better understanding and better practices regarding the use of tests for high stakes purposes, the

OCR is developing a resource guide for educators and policymakers. The guide will describe the existing non-discrimination, educational and test measurement foundations relating to the use of standardized tests that confer educational benefits to students.

Educational stakeholders at all levels have come to us requesting advice and technical assistance in a variety of test use contexts, particularly as states and districts increasingly use tests as part of their standards based reforms. And, we are addressing testing issues in a broader and more extensive array of complaints of discrimination filed with our office. These corresponding developments confirm the need to provide a useful resource that will capture legal and educational principles, references and resources to assist educators and policymakers.

We have worked with literally dozens of educator, parent, teacher, business, policymaker and testing groups and individuals to solicit input and advice regarding the scope, framing and kinds of resources to include in the guide. Notably, we have contracted with the National Academy of Sciences Board on Testing and Assessment, which has independently reviewed and which will again assess this resource to ensure that it comports with professional standards.

Perhaps the controversy that has surfaced in the wake of our extensive outreach was unavoidable. Nonetheless, we are perplexed that some of the central principles were considered by a few individuals to be novel and that the document was read by certain individuals to be something that it is not. Our effort should be understood in the clearest of terms. It is not our aim to establish standards or definitions of merit for educational institutions, as some have claimed. Nor is our effort to advocate for the elimination of standardized tests, such as the SAT. The draft of the guide reaffirms this point, in one instance stating: “[h]igh quality assessments can make high standards meaningful.” Instead, our effort is straightforward: to

explain existing legal and test measurement principles and to provide a collection of related resources—all in an effort to promote accurate decision making affecting the educational opportunities for all of our students.

Identifying and Debunking Some Myths

The misperceptions and erroneous understandings of test use principles that led to some of the controversy that has surfaced in recent weeks calls to mind the admonition by Nancy Cole, the president of the Educational Testing Service. Presciently, she has reminded us with regard to testing policies and practices that “we must acknowledge the myths that seem[] to make the issues simpler...even though acknowledging these myths makes finding solutions even more difficult.” (Cole, Merit and Opportunity: Testing and Higher Education at the Vortex, 1997). Given some of the inaccuracies regarding our ongoing work, it is an appropriate time to acknowledge—and rebut—some of the myths regarding federal non-discrimination standards and principles of sound test use.

Myth One: The goals of excellence and equity are
irreconcilable.

FALSE, as a matter of law and policy.

The view that the goals of establishing standards (as in, for instance, establishing a standard of merit in college admissions) and complying with federal laws designed to ensure non-discrimination are inconsistent is, simply, erroneous. Indeed, if the federal courts teach us anything it is this: compliance with federal non-discrimination standards rests, in the first instance upon the school’s educational judgments, to which deference is appropriately given. Correspondingly, the ultimate question upon which the federal legal analysis affecting the use of high-stakes tests depends is one of educational sufficiency: is the test valid for the purposes used? Are the inferences derived from test scores, and the educational judgments based on those inferences, accurate and fair?

The educational foundations that guide any federal legal analysis suggest that policies promoting excellence can be and should be fully aligned with the promotion of equal opportunity for all students. For the hope of a high standards education for all students to become a reality for this generation of test-taking students, we must insist on high standards for tests that have consequences for students—just as we do for schools, teachers, and the students that they teach. As foundations for the judgments that shape the lives—and lifetimes—of students, these tests must be used in ways that accurately reflect educational standards and that do not inappropriately deny opportunities to students based on their race, national origin or sex.

Myth Two: Significant disparities in the test performance by subgroups of students indicate that the test discriminates illegally.

FALSE, as a matter of law.

Test results indicating that groups of students perform differently should be a cause for further inquiry and examination, with a focus upon the relevant educational programs and testing practices at issue. The existence of significant disparities does not mean, however, that the test illegally discriminates. Differences in test scores may result from a range of factors, including: lack of preparation; poor skills or knowledge; inadequate exposure to the material tested; poor motivation; or problems with the test itself.

The guarantee under federal law is for equal opportunity—not equal results. The legal non-discrimination inquiry regarding neutral practices (referred to by the courts as the “disparate impact” standard) illustrates this point: If the educational decisions based upon test scores reflect significant disparities in the kinds of educational benefits afforded to students based on race, national origin or gender, then ask more probing questions about what’s going on to ensure non-discriminatory,

educationally sound practices. This common sense framework is paralleled in the Code of Fair Testing Practices in Education. The Code provides, in relevant part:

“Test users should...[r]eview the performance of test takers of different races, gender, and ethnic backgrounds when samples of sufficient size are available [and e]valuate the extent to which performance differences may have been caused by inappropriate characteristics of the test.” [Joint Committee on Testing Practices, Code of Fair Testing Practices in Education (1988)]

The alignment of testing principles and legal standards could not, therefore, be clearer.

Myth Three: Test scores, alone, tell the whole story.
FALSE, as a matter of good educational practice.

Tests provide very valuable guidance in making educational judgments affecting students. Decisions such as college admissions decisions frequently—and appropriately—include consideration of test scores. The value that test results can provide when making educational decisions about students does not mean, however, that test scores should as a matter of good educational practice trump the need for thoughtful educational decision making. (Note here that federal non-discrimination laws do not preclude the prospect of the permissible use of a standardized test as a sole criterion where that test has been validated for such use.)

Moreover, a test's value as an educational tool is dependent upon its design, the context in which the test is administered, and the ultimate uses of the test. For example, the SAT may be valid as a tool to be used in a university's admissions decisions. At the same time, that same test is clearly inappropriate as a basis for making decisions about whether to promote a student from eleventh to twelfth grade in high school or whether to confer a passing grade in chemistry for the year.

Even when a test is used for the purposes consistent with its design, a test is one tool among many. Just as tests are not perfect barometers of learning, conclusions based on those test results are not always error free. Many variables can affect a student's test performance, including: the quality of the student's education; the student's skill, ability, or knowledge about a particular topic; preparation for the test; or what the student ate for breakfast on the day the test was administered. Does this mean that we should do away with tests? Absolutely not. What it does suggest is precisely what test measurement standards affirm: the importance of considering multiple and educationally appropriate measures when making life-defining decisions about students. The 1985 American Psychological Association Standards for Educational and Psychological Testing state, for instance: "In elementary and secondary education, a decision ...that will have a major impact on a test taker should not automatically be made on the basis of a single test score" (APA Standard 8.12). About this point, the guidance from test developers in higher education is instructive. Consider, for instance:

- Test uses "that should be avoided" include "using test scores as the sole basis for important decisions affecting the lives of individuals, when other information of equal or greater relevance and the resources for using such information are available." [The College Board, Guidelines on the Uses of College Board Test Scores and Related Data (1988)].
- The SAT works "very well in many different circumstances...[but] there are differences in how it works for different groups of students, for different types of educational programs, and for different institutions." [The College Board, Research Notes, RN-01 (June 1997)].

Conclusion

Ultimately, good educational practices—frequently reflected in test measurement standards—and federal case law highlight the importance of considering objective measures such as tests in appropriate ways when making decisions about students. In short, they affirm that not all tests are created equal and that tests should be used in ways that are valid for the particular purpose for which they are used.

This is the driving force behind the U.S. Department of Education Office for Civil Rights' continuing effort to provide assistance to policymakers and educators as we continue to enforce federal laws that prohibit discrimination against students. Rather than creating false and polarizing "win-lose" choices on this all-important set of issues, we need to, as Secretary Riley admonishes, "step back, lower our voices, truly listen to each other and search for common ground." That is our objective as we work to fulfill the promise of longstanding education goals and non-discrimination protections: high standards learning for all students.