
nature and importance of the educational institution's goals, the educational consequences to students, the relationship of the educational institution to the student, and other factors bearing on test use, such as whether and how additional information beyond the test score enters into the educational decision at stake.¹⁹⁴ Where a test is used for promotion or graduation purposes, a major consideration is the extent to which the educational institution has provided the student with the opportunity to learn the content and skills being tested.¹⁹⁵

3. Determining Whether There Are Equally Effective Alternatives that Serve the Institution's Educational Goal with Less Disparity

If the educational institution provides sufficient evidence that the test use in question is justified educationally, the party challenging the test has the opportunity to show that there exists an equally effective alternative practice that meets the institution's goals with less disparity.¹⁹⁶ The feasibility of an alternative, including costs and administrative

at 1530 (requiring minimum ACT score for admission to undergraduate teacher education programs violated the Title VI regulations since ACT scores had not been validated for this purpose); *Sharif*, 709 F. Supp. at 361-63 (in ruling on a motion for preliminary injunction, court found that the state's use of SAT scores as the sole basis for decisions awarding college scholarships intended to reward high school achievement was not educationally justified for this purpose in that the SAT had been designed as an aptitude test to predict college success and was not designed or validated to measure past high school achievement).

Psychometric or scientific evidence is not the only way that validity can be demonstrated, however. Courts can draw inferences of validity from a wide range of data points. See *Watson*, 487 U.S. at 998 (referring to procedures used to evaluate personal qualities of candidates for managerial jobs).

¹⁹⁴ See, e.g., *Larry P.*, 793 F.2d at 980; *Georgia State Conf.*, 775 F.2d at 1417-20; *Groves*, 776 F. Supp. at 1530-31. In the educational context, tests play a complex role that bears on evaluation of educational justification. As noted by the court in *Larry P.*,

[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.

793 F.2d at 980 (quoting *Larry P. v. Riles*, 495 F. Supp. 926, 969 (1979)). Because determining whether a test is a valid basis for classifying students and placing them in different educational programs may be even more complex and difficult than determining if a test validly predicts job performance, particular sensitivity is needed to all of the interests involved. The question may be not only whether a test provides valid information about a student's ability and achievement, but whether the educational services provided to the student as a consequence of the test serve the student's needs. Inequality in the services provided to students prior to the test, as well as in the services provided as a consequence of the test, may also be a factor considered as part of the educational justification for using a test in a particular way. See *Debra P.*, 644 F.2d at 407-08 (agreeing with the statement that Title VI would not be violated if the test were a fair test of what students were taught); *Debra P. v. Turlington*, 730 F.2d 1405, 1407, 1410-11, 1416 (11th Cir. 1984) (affirming that the extent of remedial efforts to address test failure is relevant to evaluation of test use).

¹⁹⁵ See *Debra P.*, 644 F.2d at 408.

burdens, is a relevant consideration.¹⁹⁷

II. Testing Of Students With Limited English Proficiency

Testing of students with limited English proficiency in the elementary and secondary education context raises a set of unique issues. To understand the obligations of states and school districts with regard to high-stakes testing of such students, it is important to understand the basic obligations of school districts and states under Title VI and related federal law that relate to language minority students who are learning English.

Title VI prohibits discrimination based on race, color, or national origin. On May 25, 1970, the United States Department of Health, Education, and Welfare's Office for Civil Rights issued a policy memorandum entitled "Identification of Discrimination and Denial of Services on the Basis of National Origin." The May 25th memorandum clarified the responsibility of school districts, under Title VI, to provide equal educational opportunity to national origin minority group students whose inability to speak and understand the English language excludes them from effective participation in any education program offered by a school district.¹⁹⁸ This memorandum was cited with approval by the Supreme Court in its decision in *Lau v. Nichols*, which held that the district's policy of teaching national origin minority group children only in English, without any special assistance, deprived them of the opportunity to benefit from the district's education program, including meeting the English language proficiency standards required by the state for a high school diploma.¹⁹⁹ The *Lau* case held that such policies are barred when they have the effect of denying such benefits, even though no purposeful design is present.²⁰⁰

Subsequently, *Castaneda v. Pickard*,²⁰¹ relying on the language of the Equal

¹⁹⁶ See *New York Urban League v. New York*, 71 F.3d 1031, 1036 (2d Cir. 1995) (stating "the plaintiff may still prove his case by demonstrating that other less discriminatory means would serve the same objective"); see also *Albemarle Paper Co. v. Moody*, 422 U.S. 405, 425 (1975); *Richardson*, 729 F. Supp. at 815.

¹⁹⁷ See *Wards Cove*, 490 U.S. at 661 (indicating that factors such as costs or other burdens are relevant in determining whether the alternative is equally effective in serving employer's legitimate goals); *MacPherson v. University of Montevallo*, 922 F.2d 766, 773 (11th Cir. 1991) (holding that plaintiff must show that the alternative is economically feasible); *Sharif*, 709 F. Supp. at 363-64 (finding defendant's claim that proposed alternative was not feasible and excessively burdensome not persuasive since most other states used proposed alternative).

¹⁹⁸ See *Identification of Discrimination and Denial of Services on the Basis of National Origin*, 35 Fed. Reg. 11595 (1970). The Department of Health, Education and Welfare was the predecessor of the U.S. Department of Education.

¹⁹⁹ See *Lau*, 414 U.S. at 566-68.

²⁰⁰ See *Lau*, 414 U.S. at 568 (citing, among other legal authority, the predecessor of 34 C.F.R. § 100.3 (b)(2)).

Educational Opportunities Act (EEOA), explained the steps school districts must take to help students with limited English proficiency overcome language barriers to ensure that they can participate meaningfully in the districts' educational programs.²⁰² The court stated that school districts have an obligation to provide services that enable students to acquire English language proficiency. A school system that chooses to temporarily emphasize English over other subjects retains an obligation to provide assistance necessary to remedy academic deficits that may have occurred in other subjects while the student was focusing on learning English.

Under the *Castaneda* standards, school districts have broad discretion in choosing a program of instruction for limited English proficient students. However, the program must be based on sound educational theory, must be adequately supported so that the program has a realistic chance of success, and must be periodically evaluated and revised, if necessary, to achieve its goals.

The disparate impact framework discussed above may also be used to examine whether tests used for high-stakes purposes result in a discriminatory impact upon students with limited English proficiency. As part of this analysis, questions may arise regarding the validity and reliability of the test for these students.²⁰³ Depending upon the purpose of the test and the characteristics of the populations being tested, in some situations, accommodations or other forms of assessment of the same construct may be necessary. In short, the obligation is to ensure that the same constructs are being measured for all students.

There are three particularly important areas involving high-stakes testing of students with limited English proficiency: (1) tests used to determine a student's proficiency in the areas of speaking, listening, reading, or writing English for the purpose of determining whether the student should be provided with a program to enable the student to acquire English language skills (and, later, for the purpose of determining whether the student is ready to exit the program); (2) tests used to determine if the student meets the criteria for other specialized instructional programs, such as gifted and talented or vocational education programs; and (3) systemwide tests, including graduation tests, administered to determine if students have met performance standards.

Tests used to determine a student's initial and continuing need for special language

²⁰¹ See *Castaneda v. Pickard*, 648 F.2d 989, 1005-06, 1009-12 (5th Cir. 1981). The analytical framework in *Castaneda* which was decided under the Equal Educational Opportunities Act (EEOA), 20 U.S.C. §§ 1701 *et seq.*, has been applied to OCR's Title VI analysis. See *Williams Memorandum*, *supra* note 45. The EEOA contains standards related to limited English proficient students similar to the Title VI regulations.

²⁰² See *Castaneda*, 648 F.2d at 1011.

²⁰³ See discussion *supra* Chapter 1 (Test Measurement Principles) Part (II)(B) (Testing of Limited English Proficient Students) for a discussion of the psychometric principles involved in determining the reliability and validity of tests used with limited English proficient students.

programs should be appropriate in light of a district's own performance expectations and otherwise valid and reliable for the purpose used. Tests used by schools to help select students for specialized instructional programs, including programs for gifted and talented students, should not screen out limited English proficient students unless the program itself requires proficiency in English for meaningful participation.²⁰⁴ When a state or school district adopts content and performance standards, and uses high-stakes tests, such as graduation tests, to measure whether students have mastered these standards, a critical factor is whether the overall educational program provided to students with limited English proficiency is reasonably calculated to enable the students to master the knowledge and skills that are required to pass the test. When education agencies institute standards-based testing, it is important for them to examine their programs for students with limited English proficiency to determine when and how these students will be provided with the instruction needed to prepare them to pass the test in question.²⁰⁵

In addition, students with limited English proficiency may not be categorically excluded from standardized testing designed to increase accountability of educational programs for effective instruction and student performance. If these students are not included, the test data will not fairly reflect the performance of all students for whom the education agency is responsible.²⁰⁶ Such test data can also help a district to assess the effectiveness of its content and English language acquisition programs.

For information on the factors that help ensure accuracy of tests for limited English proficient students, see discussion *infra* Chapter 1 (Test Measurement Principles) Part II (B) (Testing of Limited English Proficient Students). In making decisions about testing limited English proficient students, factors such as the student's level of English proficiency, the primary language of instruction, the level of literacy in the native language, and the number of years of instruction in English may all be pertinent.¹

²⁰⁴ See *Williams Memorandum*, *supra*, note 45.

²⁰⁵ Careful attention to the alignment between instructional content and testing standards is especially important for students who receive instruction that deviates from the regular curriculum. See *Brookhart v. Illinois State Bd. of Educ.*, 697 F.2d 179, 186-87 (7th Cir. 1982) (students with disabilities in special education programs denied exposure to material covered in newly instituted graduation test).

²⁰⁶ Indeed, Title I of the Elementary and Secondary Education Act explicitly requires states to include limited English proficient students in the statewide assessments used to hold schools and school districts accountable for student performance. See Title I of the Elementary and Secondary Education Act, 20 U.S.C. § 6311(b)(3)(F)(iii). If a school district uses the results of a test given for program accountability purposes to make educational decisions about individual students, the high-stakes use of the test must also be valid and reliable for this purpose.

When students participate in assessments designed to meet the requirements of Title I of the Elementary and Secondary Education Act, as amended, those assessments must be implemented in a manner that is consistent with both the requirements of Title VI and Title I.

III. Testing Of Students With Disabilities

Three federal statutes provide basic protections for elementary and secondary students with disabilities. Section 504 of the Rehabilitation Act of 1973 (Section 504) and Title II of the Americans with Disabilities Act of 1990 (Title II) prohibit discrimination against persons with disabilities by public schools.²⁰⁸ The Individuals with Disabilities Education Act (IDEA) establishes rights and protections for students with disabilities and their families. It also provides federal funds to state education agencies and school districts to assist in educating students with disabilities.²⁰⁹ Under Section 504, Title II, and the IDEA,²¹⁰ school districts have a responsibility to provide students with disabilities, as defined by applicable law, with a free appropriate public education. Providing effective instruction in the general curriculum for students with disabilities is an important aspect of providing a free appropriate public education.

The regulations implementing Section 504 and Title II specifically provide that a recipient of federal funds may not “utilize criteria or methods of administration ... that have the effect of subjecting qualified persons with disabilities to discrimination on the basis of disability.”²¹¹ Under Section 504, Title II, and the IDEA, tests given to students with disabilities must be selected and administered so that the test accurately reflects what a student knows or is able to do, rather than a student’s disability (except when the test is designed to measure disability-related skills). This means that students with

²⁰⁷ For more information on appropriate ways of testing students who are learning English, see Rebecca J. Kopriva, Council of Chief State School Officers, *Ensuring Accuracy in Testing for English Language Learners* (2000).

²⁰⁸ Although this part of the chapter deals only with students with disabilities attending public elementary and secondary schools, private schools that are not religious schools operated by religious organizations are covered by Title III of the ADA. Title III of the Americans with Disabilities Act of 1990, 42 U.S.C. §§ 12181 *et seq.* In addition, Title I of the Elementary and Secondary Education Act of 1965, as amended, contains important provisions regarding students with disabilities in the Title I program and their participation in assessments of Title I programs. See 20 U.S.C. § 6311(b)(3)(F).

²⁰⁹ See Individuals with Disabilities Education Act, 20 U.S.C. § 1400(d)(1)(c).

²¹⁰ The Section 504 regulation is found at 34 C.F.R. Part 104. The Title II regulation is found at 28 C.F.R. Part 35. The IDEA regulation is found at 34 C.F.R. Part 300.

²¹¹ See the Title II regulation at 28 C.F.R. 35.130(b)(3) and the Section 504 regulation at 34 C.F.R. § 104.4(b)(4). In *Guardians Association*, the United States Supreme Court upheld the use of the effects test in the context of Title VI, stating that the Title VI regulation forbids the use of federal funds, “not only in programs that intentionally discriminate on racial grounds but also in those endeavors that have a [racially disproportionate] impact on racial minorities.” 463 U.S. at 589.

disabilities covered by these statutes must be given appropriate accommodations and modifications in the administration of the tests that allow the same constructs to be measured. Examples include oral testing, tests, , Braille versions of tests, individual testing, and separate group testing.

Generally, there are three critical areas in which high-stakes testing issues arise for students with disabilities: (1) tests used to determine whether a student has a disability and, if so, the nature of the disability; (2) tests used to determine if the student meets the criteria for other specialized instructional programs, such as gifted and talented or vocational education programs; and (3) systemwide tests administered to determine if a student has met performance standards.

Under Section 504, Title II, and the IDEA, before an elementary and secondary school student can be classified as having a disability, the responsible education agency must individually evaluate the student in accordance with specific statutory and regulatory requirements, including requirements regarding the validity of tests and the provision of appropriate accommodations.²¹² These requirements prohibit the use of a single test score as the sole criterion for determining whether a student has a disability and for determining an appropriate educational placement for the student.²¹³

When tests are used for other purposes, such as in making decisions about placement in gifted and talented programs, it is important that tests measure the skills and abilities needed in the program, rather than the disability, unless the test purports to measure skills or functions that are impaired by the disability and such functions are necessary for participation in the program.²¹⁴ For this reason, appropriate accommodations may need to be provided to students with disabilities in order to measure accurately their performance in the skills and abilities required in the program.

Furthermore, federal nondiscrimination laws would generally require the inclusion of students with disabilities in state- and districtwide assessment programs, including high-stakes tests, except as participation in particular tests is individually determined to be inappropriate for a particular student. Such assessments provide valuable information which benefits students, either directly, such as in the measurement of individual progress against standards, or indirectly, such as in evaluating programs. Given these benefits, exclusion from assessment programs, unless such participation is individually

²¹² See 34 C.F.R. § 104.35(b) (specific provisions covering the use of tests for evaluation purposes).

²¹³ See 34 C.F.R. § 104.35(c) (requiring placement decisions to consider information from a variety of sources).

²¹⁴ See 34 C.F.R. §§ 104.35(b)(3), 300.532.

determined inappropriate because of the student's disability, would generally violate Section 504 and Title II. If a student with a disability will take the systemwide assessment test, including a high-stakes test, the student must be provided appropriate instruction and appropriate test accommodations.²¹⁵

In addition, the Individuals with Disabilities Education Amendments of 1997 specifically require states, as a condition of receiving IDEA funds, to include students with disabilities in the regular state- and districtwide assessment programs, with appropriate accommodations, where necessary.²¹⁶ The IDEA requirements cover tests with high-stakes consequences given to measure individual achievement as well as tests given for program accountability purposes. The IDEA also requires state or local education agencies to develop guidelines for the relatively small number of students with disabilities who cannot take part in state- and districtwide tests to participate in alternate assessments.²¹⁷

For children with disabilities, school personnel knowledgeable about the student, the nature of the disability, and the testing program, in conjunction with the student's parent or guardian, determine whether the student will participate in all or part of the state or districtwide assessment of student achievement.²¹⁸ The decision must be documented in the student's individualized education program (IEP), or a similar record, such as a Section 504 plan. These records must also state any individual accommodations in the administration of the state- or districtwide assessments of student achievement that are needed to enable the student to participate in such assessment. An IEP, developed under the IDEA, must also explain how the student will be assessed if it is inappropriate for the student to participate in the testing program even with accommodations.²¹⁹ The individual decisions made regarding testing of the student in the IEP or Section 504 plan are subject to appeal by the parent or guardian through the due process procedures required by applicable law.²²⁰

²¹⁵ See *Brookhart*, 697 F.2d at 183-84. Some courts have held that a student with a disability may be denied a diploma if, despite receiving appropriate services and testing accommodations, the student, because of the disability, is unable to pass the required test or meet other graduation requirements. See *id.*, at 183; *Anderson, v. Barks* 520 F. Supp. 472, 509-11 (sd. 6a 1991); *Board of Educ. v. Ambach*, 458 N.Y.S.2d 680, 684-85, 689 (N.Y. App. Div. 1982), *aff'd*, 469 N.Y.S.2d 669 (1983).

²¹⁶ See 20 U.S.C. § 1412(a)(17); 34 C.F.R. § 300.138(a).

²¹⁷ See 34 C.F.R. § 300.138(b). The IDEA Final Regulations, Attachment I—Analysis of Comments and Changes, 64 Fed. Reg. 12406, 12564 (1999), projects that there will be a relatively small number of students who will not be able to participate in the district or state assessment program with accommodations and modifications, and will therefore need to be assessed through alternate means. These alternate assessments must be developed and conducted beginning not later than July 1, 2000.

²¹⁸ See 34 C.F.R. § 300.347(a)(5) (IEP requirements applicable to assessment of students with disabilities under IDEA); 34 C.F.R. § 104.33 (more general evaluation requirements under Section 504).

²¹⁹ See 34 C.F.R. § 300.347(a)(5).

²²⁰ See 34 C.F.R. §§ 300.507, 104.36.

Section 504 and Title II also prohibit discrimination against qualified persons with disabilities in virtually all public and private post-secondary institutions.²²¹ The regulatory requirements related to disability discrimination are different in post-secondary education than in elementary and secondary education. Post-secondary institutions are not required to evaluate students or to provide them with a free appropriate education.

High-stakes testing issues at the post-secondary level generally relate to tests considered by post-secondary institutions for admissions, including tests given by an educational institution or other covered entities as prerequisites for entering a career or career path, and tests of academic competency required by the institution to complete a program. This guide is not intended to offer a complete or detailed explanation of each of these testing situations, but only a brief synopsis.²²²

The Section 504 regulation specifically provides that higher education institutions' admissions procedures may not make use of any test or criterion for admission that has a disproportionate, adverse impact on individuals with disabilities unless (1) the test or criterion, as used by the institution, has been validated as a predictor of success in the education program or activity and (2) alternative tests or criteria that have a less disproportionate, adverse impact are not shown to be available.²²³ In administering tests, appropriate accommodations must be provided so that the person can demonstrate his or her aptitude and achievement, not the effect of the disability (except where the functions impaired by the disability are the factors the test purports to measure).²²⁴

For other high-stakes tests that an institution might administer, such as rising junior

²²¹ Under the Section 504 regulation, a qualified person with a disability for purposes of post-secondary education is an individual with a disability within the meaning of the regulation who meets the academic and technical standards for admission. See 34 C.F.R. §§ 104.3(j), 104.3(k).

²²² Test providers that are not higher education institutions may be covered by Section 504 if they receive federal funds; by Title II if they are parts of governmental units; or by Title III if they are private entities. Each of these laws has its own requirements. For more information regarding testing under Title III of the ADA, consult the U.S. Department of Justice.

²²³ See 34 C.F.R. § 104.42(b)(2). Appendix A to the Section 504 regulation, Subpart E-Post-secondary Education, No. 29, notes that the party challenging the test would have the burden of showing that alternate tests with less disparate impact are available.

²²⁴ See 34 C.F.R. § 104.42(b)(3).

²²⁵ Some undergraduate college programs require students to pass a rising junior examination to determine whether students have met the college's standards in writing or other academic skills as a prerequisite for advancement to junior year status.

tests, similar requirements apply.²²⁵ The institution must provide adjustments or accommodations and auxiliary aids and services that enable the student to demonstrate the knowledge and skills being tested.²²⁶

Students are required to notify the educational institution when accommodations are needed and supply adequate documentation of a current disability and the need for accommodation.²²⁷ The student's preferred accommodation does not have to be provided as long as an effective accommodation is provided.

Test accommodations are intended to provide the person with disabilities the means by which to demonstrate the skills and knowledge being tested. Although Section 504 and Title II require a college or university to make reasonable modifications, neither Section 504 nor Title II requires a college or university to change, lower, waive, or eliminate academic requirements or technical standards that can be demonstrated by the college or university to be essential to its program of instruction or to any directly related licensing requirement.²²⁸ Accommodations requested by students need not be provided if they would result in a fundamental alteration to the institution's program.²²⁹

IV. constitutional Protections

In addition to applying federal nondiscrimination statutes, courts have also considered constitutional issues that may arise when public school districts or state education agencies utilize tests for high-stakes purposes in their educational programs, particularly tests required for promotion or graduation.²³⁰ constitutional challenges to testing programs under the Fourteenth Amendment have raised both equal protection and due process claims. The equal protection principles involved in discrimination cases are, generally speaking, the same as the standards applied to intentional discrimination claims under the applicable federal nondiscrimination statutes.²³¹

²²⁶ See 34 C.F.R. §§ 104.44(a), 104.44(d).

²²⁷ See, e.g., *Kaltenberger v. Ohio College of Podiatric Medicine*, 162 F.3d 432, 437 (6th Cir. 1998).

²²⁸ See 34 C.F.R. § 104.44(a).

²²⁹ See *Southeastern Community College v. Davis*, 442 U.S. 397, 413 (1979); *Wynne v. Tufts Univ. Sch. of Medicine*, 976 F.2d 791, 794-96 (1st Cir. 1992), *cert. denied*, 507 U.S. 1030 (1993).

²³⁰ The U.S. Department of Education, Office for Civil Rights, does not have jurisdiction to resolve constitutional cases. However, some cases involve constitutional issues that overlap with discrimination issues arising under federal civil rights laws.

²³¹ Federal cases may also involve equal protection challenges to a jurisdiction's use of tests in which the claim is not based on race or sex discrimination, but, instead, on allegations that flaws in the jurisdiction's use of a test make the resulting educational decisions arbitrary and capricious or fundamentally unfair. See *GI Forum*, 87 F. Supp. 2d at 682. As a general matter, courts express reluctance to second guess a state's educational policy choices when faced with such challenges, although recognize that a state cannot "exercise that [plenary] power without reason and without regard to the United States Constitution." *Debra P.*, 644 F.2d at 403. When there is

The due process clause of the Fourteenth Amendment is particularly associated with cases challenging the adequacy of the notice provided to students prior to this type of test and the students' opportunity to learn the required content.²³² In analyzing such due process claims, courts have generally considered three issues:

- (1) Is the testing program reasonably related to a legitimate educational purpose?

Federal courts typically defer to educators' policy judgments regarding the value of the educational benefits sought from the testing programs.²³³ For example, improving the quality of elementary and secondary education through the establishment of academic

no claim of discrimination based on membership in a suspect class, the equal protection claim is reviewed under the rational basis standard. In these cases, the jurisdiction need show only that the use of the tests has a rational relationship to a valid state interest. See *id.*, at 406; see also *Erik V. v. Causby*, 977 F. Supp. 384, 389 (E.D.N.C. 1997).

²³² A review of relevant cases reveals the highly fact and context-specific nature of the conclusions reached by federal courts considering alleged violations of the due process clause. In *Debra P.*, the Fifth Circuit held that students' due process rights were violated when a newly imposed minimum competency test required for high school graduation was instituted without adequate notice and an opportunity for students to learn the material covered by the test. See *Debra P.*, 644 F.2d at 404. Three years later, in *Debra P. v. Turlington*, the court held that students who now had six years notice of the exam were afforded the opportunity to learn the relevant material, given the state's remedial programs. See *Debra P.*, 730 F.2d at 1416-17. For additional courts identifying due process violations in the way in which a competency test was instituted, see *Brookhart*, 697 F.2d at 186-87 (holding that district-required minimum competency test for graduation denied due process to students with disabilities where notice was inadequate and students had not been exposed to 90 percent of the material covered by the test); *Crump v. Gilmer Indep. Sch. Dist.*, 797 F. Supp. 552, 556-57 (E.D. Tex. 1992) (granting temporary restraining order where district had not demonstrated validity of graduation examination in light of actual instructional content); *Anderson v. Banks*, 520 F. Supp. 472, 508-09 (S.D. Ga. 1981) (finding that school district failed to show that minimum competency test required for high school graduation covered material actually taught at school). Other cases have concluded that adequate notice was provided, the test or criterion at issue was closely related to the instructional program, or the promotion decision was not shown to be outside the discretion of school authorities. See *Erik V.*, 977 F. Supp. at 389-90 (finding that promotion decision was within proper purview of school authorities); *Williams v. Austin Indep. Sch. Dist.*, 796 F. Supp. 251, 253-54 (W.D. Tex. 1992) (considering students to have had seven years advance notice of high school competency exam although standards of performance were recently raised). Also relevant are promotion cases in which students were required to demonstrate adequate reading skills, although a separate test was not apparently involved. See *Bester v. Tuscaloosa City Bd. of Educ.*, 722 F.2d 1514, 1516 (11th Cir. 1984) (finding reading standards required for promotion to merely reinforce district policy of retention for substandard work); *Sandlin v. Johnson*, 643 F.2d 1027, 1029 (4th Cir. 1981) (finding denial of second-grade promotion for failing to attain required level in reading series within discretion of school district). For a testing case raising similar due process issues at the post-secondary level, see *Mahavongsanan v. Hall*, 529 F.2d 448, 450 (5th Cir. 1976) (finding no violation of due process where the university's decision to require a comprehensive examination for receipt of a graduate degree was a reasonable academic regulation, plaintiff received timely notice that she would be required to take the examination, she was allowed to retake the test, and the university afforded her an opportunity to complete additional course work in lieu of the examination).

²³³ See *Regents of the Univ. of Mich. v. Ewing*, 474 U.S. 214, 226-27 (1985); *Debra P.*, 644 F.2d at 406; *Anderson*, 520 F. Supp. at 506.

standards has been seen as a legitimate goal of a testing program, and colleges and universities generally have been given wide latitude in framing degree requirements and making academic decisions.²³⁴ The constitutional inquiry goes on to examine whether the challenged testing program is reasonably related to the educators' legitimate goals or whether the program produces results that are arbitrary and capricious or fundamentally unfair.²³⁵

(2) Have students received adequate notice of the test and its consequences?

In the elementary and secondary school context, courts have required sufficient advance notice of tests required for graduation to give students a reasonable chance to learn the material presented on the test.²³⁶ A particularly important concern in some of these decisions is the adequacy of notice provided to students. This issue has arisen in cases where racial minority students and students with disabilities received inadequate notice and did not receive a program of instruction that prepared them to pass the test.²³⁷ In looking at the length of the transition period needed between the announcement of a new requirement and its full implementation, the kind of test and the context in which it is administered are central factors to be considered. Specific circumstances taken into account include the nature of instructional supports, including

²³⁴ See *Ewing*, 474 U.S. at 222, 226-27. (acknowledging that courts will not review academic decisions of colleges and universities unless the decision is such a substantial departure from accepted academic norms as to demonstrate that professional judgment was not actually exercised or where discrimination is claimed); *Debra P.*, 644 F.2d at 402 (finding praiseworthy a state's effort to set standards to improve public education).

²³⁵ The determination as to whether a testing program is rationally related to a legitimate educational goal is technically considered as one of substantive due process under the Fourteenth Amendment. See *Debra P.*, 644 F.2d at 406; *Anderson*, 520 F. Supp. at 506. As discussed in the Introduction to this guide, where a test is used as a graduation or promotion requirement, the opportunity to learn the content tested is generally part of this determination. See *GI Forum*, 87 F. Supp. 2d at 682. Insofar as due process cases may involve other technical questions of the validity of the test used to address the institution's goal, these issues are discussed in the portions of the guide addressing discrimination under federal civil rights laws.

²³⁶ Although there are important exceptions, (*United States v. LULAC*, 793 F.2d 636, 648 (5th Cir. 1986), and *Anderson*, 520 F. Supp. at 505), courts have often considered the issue of adequate notice to be one of procedural due process. For procedural due process to apply, a protected property or liberty interest must be identified. See *Brookhart*, 697 F.2d at 185 (identifying a liberty interest, based on stigma of diploma denial, that disastrously affected plaintiffs' future employment and educational opportunities); *Debra P.*, 644 F.2d at 404 (finding sufficient to trigger due process protection a state-created mutual expectation that students who successfully complete required courses would receive diploma); *Erik V.*, 977 F. Supp. at 389-90 (finding no property interest in grade-level promotion warranting preliminary injunction).

²³⁷ See *Brookhart*, 697 F.2d at 186-88; *Debra P.*, 644 F.2d at 404.

²³⁸ See *Debra P.*, 730 F.2d at 1407, 1410-12, 1415-16; *Anderson*, 520 F. Supp. at 505.

²³⁹ Re-testing was available in *Erik V.*, 977 F. Supp. at 388-89, and in *Anderson*, 520 F. Supp. at 505.

remediation, that accompany the test,²³⁸ whether re-testing is permitted,²³⁹ and whether the decision to promote or graduate the student considers other information about the student's performance.²⁴⁰

(3) Are students actually taught the knowledge and skills measured by the test?

Several courts have found that “fundamental fairness” requires that students be taught the material covered by the test where passing the test is a condition for receipt of a high school diploma.²⁴¹ For example, in analyzing this issue in a case involving a state where there had been past intentional segregation in elementary and secondary schools before a statewide diploma test was required, and where racial minority students had a disproportionate failure rate on the test, the court took the state's past intentional segregation into account in determining whether racial minority students had had adequate opportunities to learn the material covered by the test.²⁴² For the test to meaningfully measure student achievement, the test, the curriculum, and classroom instruction should be aligned. In cases examining systemwide administration of a test, courts require evidence that the content covered by the test is actually taught, but may not expect proof that every student has received the relevant instruction.²⁴³

²⁴⁰ See *Erik V.*, 977 F. Supp. at 387 (reading performance of students with grades of A, B, or C on grade-level work was further reviewed by teacher and principal to determine if student should be promoted notwithstanding the failing test score).

²⁴¹ The question of instructional or curricular validity is usually posed as one of substantive due process. See *Brookhart*, 697 F.2d at 184-87; *Debra P.*, 644 F.2d at 406; *Anderson*, 520 F. Supp. at 509.

²⁴² See *Debra P.*, 644 F.2d at 407 (where black students disproportionately failed a statewide test necessary to obtain a high school diploma, and, due to the prior dual school system, black students received a portion of their education in unequal, inferior segregated schools, and where the state was unable to show that the diploma sanction did not perpetuate the effects of that past intentional discrimination, the court found that immediate use of the diploma sanction punished the black students for deficiencies created by the dual school system in violation of their constitutional right to equal protection); *Debra P.*, 474 F. Supp. 244, 257 (M.D. Fla. 1979) (“punishing the victims of past discrimination for deficits created by an inferior educational environment neither constitutes a remedy nor created better educational opportunities”).

²⁴³ See *Anderson, v. Banks* 540 F. Supp. 761, 765 (S.D. Ga. 1982).

APPENDIX A: Glossary of Legal Terms

This glossary is provided as a plain language reference to assist non-lawyers in understanding commonly used legal terms that are either used in this guide or are important to know in understanding the terms in the guide. Legal terms are often “terms of art.” In other words, they mean something slightly different or more specific in the legal context than they do in ordinary conversation.

Burden of proof—the duty of a party to substantiate its claim or defense against the other party. In civil actions, the weight of this proof is usually described as a preponderance of the evidence. See BLACK’S LAW DICTIONARY 196-97 (6th ed. 1990); see also *Disparate Impact*.

constitutional rights—the rights of each American citizen that are guaranteed by the United States Constitution. See *Brown v. Board of Educ.*, 347 U.S. 483 (1954); *Bolling v. Sharpe*, 347 U.S. 497 (1954); BLACK’S LAW DICTIONARY 312 (6th ed. 1990).

De jure segregation or discrimination— term applied to systemic school segregation that was mandated by statute or that was accomplished through the intentionally segregative actions of local school districts or state agencies.

Different treatment—a claim that similarly situated persons are treated differently because of their race, color, national origin, sex or disability. Under federal non-discrimination laws, policies and practices must be applied consistently to an individual or group of students regardless of their race, national origin, sex, or disability, unless there is a lawful reason for not doing so. To prove different treatment, one must show that “a challenged action was motivated by an intent to discriminate.” *Elston v. Talladega County Bd. of Educ.*, 997 F.2d 1394, 1406 (11th Cir. 1993). This requires a showing that the decisionmaker was not only aware of the person’s race, national origin, sex, or disability, but that the recipient acted, at least in part, because of the person’s race, national origin, sex or disability. However, the record need not contain “direct evidence of bad faith, ill will or any evil motive,” on the part of the recipient. See *Elston*, 997 F.2d at 1406 (quoting *Williams v. City of Dotham*, 745 F.2d 1406, 1414 (11th Cir. 1984)). Evidence of discriminatory intent may be direct or circumstantial. Different treatment may be justified by a lawful reason, for example, to remedy prior discrimination. See generally *United States v. Fordice*, 505 U.S. 717, 728-30 (1992); *Wygant v. Jackson Bd. of Educ.*, 476 U.S. 267, 290-91 (1986); *Regents of the Univ. of Cal. v. Bakke*, 438 U.S. 265, 305-20 (1978); *Hopwood v. Texas*, 78 F.3d 932, 948-50 (5th Cir. 1996), cert. denied, 518 U.S. 1033 (1996); BLACK’S LAW DICTIONARY 470 (6th ed. 1990).

Disparate impact—disparate impact analysis applies when the application of a neutral criterion or a facially neutral practice has discriminatory effects and the criterion or practice is not determined to be “educationally justified” or “educationally necessary.” In contrast to intentional discrimination, the disparate impact analysis does not require proof of discriminatory motive. Under the disparate impact analysis, the party challenging the criterion or practice has the burden of establishing disparate impact. If disparate impact is established, the party defending the practice must establish an “educational justification.” If the educational institution provides sufficient evidence that the test use in question is justified educationally, the party challenging the test has the opportunity to show that there exists an alternative practice that meets the institution’s goals as well as the challenged test use and that would eliminate or reduce the adverse impact. See *Board of Educ. v. Harris*, 444 U.S. 130, 143 (1979); *Georgia State Conf. of Branches of NAACP v. Georgia*, 775 F.2d 1403, 1412 (11th Cir. 1985); *Groves v. Alabama State Bd. of Educ.*, 776 F. Supp. 1518 (M.D. Ala. 1991),

Dual system—a previously segregated educational system in which black and white schools, ostensibly similar, existed side-by-side. See *Brown v. Board of Educ.*, 347 U.S. 483 (1954); *Anderson v. Banks*, 520 F. Supp. 472, 499-501 (S.D. Ga. 1981).

Due process—a constitutionally guaranteed right. The Fifth Amendment states that no citizen shall “be deprived of life, liberty, or property, without due process of law.” The Fourteenth Amendment applied this passage to the states as well. Today it is used by the judiciary to define the scope of fundamental fairness due to each citizen in his or her interactions with the government and its agencies. Some courts have held that a student’s expectation in receiving a high school diploma in return for meeting certain attendance and academic criteria is a form of a property right or liberty interest. See *Debra P. v. Turlington*, 644 F.2d 397 (5th Cir. 1981); *Crump v. Gilmer Indep. Sch. Dist.*, 797 F. Supp. 552, 555-56 (E.D. Tex. 1992). But see *Board of Educ. v. Ambach*, 458 N.Y.S.2d 680, (N.Y. App. Div. 1982), *aff’d*, 457 N.E.2d 775 (1983); BLACK’S LAW DICTIONARY 500-01 (6th ed. 1990); see also *Procedural Due Process*, *Substantive Due Process*.

Educational necessity—once the party challenging the practice has shown a significant disparate impact, the educational institution using the challenged practice must present sufficient evidence that it is justified by educational necessity. Educational necessity generally refers to a showing that practices or procedures are necessary to meeting an important educational goal. See *Elston v. Talladega County Bd. of Educ.*, 997 F.2d 1394, 1412 (11th Cir. 1993) (citing *Georgia State Conf. of Branches of NAACP v. Georgia*, 775 F.2d 1403, 1412, 1417 (11th Cir. 1985)). In the context of testing this means the test or assessment procedure must serve a legitimate educational goal and be valid and reliable for the purpose used.

Equal protection—classifications based on race, sex or other grounds may be challenged under the equal protection clause of the Fourteenth Amendment to the U.S. Constitution when imposed by state or local government agencies. Distinctions explicitly based on race or ethnicity, neutral criteria having a discriminatory purpose or other intentionally discriminatory conduct based on race or ethnicity will violate the Fourteenth Amendment, unless the action is narrowly tailored to serve a compelling purpose. Intentional sex discrimination will violate the Fourteenth Amendment unless there is an exceedingly persuasive justification. See *United States v. Virginia*, 518 U.S. 515 (1996). Distinctions based on other grounds will not violate the equal protection clause unless they are not rationally related to a legitimate governmental objective.

Facially neutral—a regulation, rule, practice or other activity that does not appear to be discriminatory. A facially neutral practice may be found in violation of federal law if the practice results in significant differences in the distribution of benefits or services to persons based on race, national origin, sex or disability without a substantial legitimate educational justification or there are equally or comparably alternative practices available that meet the institution's goals with less disparate impact. See *Lau v. Nichols*, 414 U.S. 563 (1974); *Larry P. v. Riles*, 793 F.2d 969 (9th Cir. 1984).

High-stakes educational decisions for students—decisions that have significant impact or consequences for individual students. These decisions may involve student placement in gifted and talented programs; decisions concerning whether a student has a disability; the appropriate educational program for a student with a disability; promotion or graduation decisions; and higher education admissions decisions and scholarship awards. See National Research Council, *High Stakes: Testing for Tracking, Promotion, and Graduation*, pp. 1-2 (Jay P. Heubert & Robert Hauser eds., 1999); *Larry P. v. Riles*, 793 F.2d 969 (9th Cir. 1984); *Sharif v. New York State Educ. Dep't*, 709 F. Supp. 345 (S.D.N.Y. 1989).

Less discriminatory alternative—if the education institution presents sufficient evidence that the test use or educational practice in question is justified educationally, the party challenging the test has the opportunity to show that there exists an equally or comparably effective alternative practice that meets the institution's goals and that would eliminate or reduce the adverse impact. See *Elston v. Talladega County Bd. of Educ.*, 997 F.2d 1394, 1407 (11th Cir. 1993); *Georgia State Conf. of Branches of NAACP v. Georgia*, 775 F.2d 1403 (11th Cir. 1985). Costs and administrative burdens are among the factors considered in assessing whether the alternative practice is equally effective in fulfilling the institution's goals. See *Wards Cove Packing Co. v. Antonio*, 490 U.S. 642, 661 (1989); *Sharif v. New York State Educ. Dep't*, 709 F. Supp. 345, 363-64 (S.D.N.Y. 1989) (defendant's claim that proposed alternative was not feasible and excessively burdensome not persuasive since most other states used proposed alternative).

Procedural due process—the right each American citizen has under the Constitution to a fair process in actions that affect an individual's life, liberty or property. Procedural due process includes notice and the right to be heard. Some courts have found that procedural due process applies to the implementation of minimum competency examinations required for high school graduation. See *Debra P. v. Turlington*, 474 F. Supp. 244, 263-64 (M.D. Fla. 1979), *aff'd in part and vacated in part*, 644 F.2d 397 (5th Cir. 1981); *Erik V. v. Causby*, 977 F. Supp. 384, 389-90 (E.D.N.C. 1997); *Crump v. Gilmer Indep. Sch. Dist.*, 797 F. Supp. 552, 555-56 (E.D. Tex. 1992); BLACK'S LAW DICTIONARY 1203 (6th ed. 1990).

Significantly disproportionate—when statistical analysis shows that the success rate of members of an identified group is significantly lower than would be expected from random distribution within the appropriate qualified pool, the test in question is said to have a disproportionate adverse impact. There is no set formula to determine when a sufficient level of adverse impact has been reached; the Supreme Court has stated that statistical disparities must be sufficiently substantial that they raise an inference of causation. Courts have advanced percentage disparities, standard deviations or other statistical formulae to address this component. Disparate impact itself does not necessarily mean that discrimination has taken place, but it does trigger an inquiry regarding the educational justification of the challenged practice. See *Watson v. Fort Worth Bank & Trust*, 487 U.S. 977, 994-95 (1988); *Groves v. Alabama State Bd. of Educ.*, 776 F. Supp. 1518, 1529-32 (M.D. Ala. 1991); *Richardson v. Lamar County Bd. of Educ.*, 729 F. Supp. 806, 815-16 (M.D. Ala. 1989), *aff'd*, 935 F.2d 1240 (11th Cir. 1991).

Statutory rights—rights protected by statute, as opposed to constitutional rights, which are protected by the Constitution.

Substantive due process—often stated as “fundamental fairness.” In an education context, proof that students had not been taught the material on which they were tested might be a substantive due process violation. Some courts have held that students have the equivalent of a property or liberty interest in graduating or being promoted according to the expectations given them. See *Debra P. v. Turlington*, 644 F.2d 397 (5th Cir. 1981); *Crump v. Gilmer Indep. Sch. Dist.*, 797 F. Supp. 552, 555-56 (E.D. Tex. 1992); BLACK'S LAW DICTIONARY 1429 (6th ed. 1990).

Unitary system—a desegregated school system. The Supreme Court has held that all previously intentionally segregated school systems are required to become unitary systems. Although the term has been interpreted in different ways by different courts, a “unitary system” is typically one in which all vestiges of past discrimination and

segregated practices have been eliminated. See *Freeman v. Pitts*, 506 U.S. 467, 486-89 (1992); *Board of Educ. v. Dowell*, 498 U.S. 237, 243-46, 249-51 (1991); *Keyes v. School Dist. No. 1*, 413 U.S. 189, 208, 257-58 (1973); *Georgia State Conf. of Branches of NAACP v. Georgia*, 775 F.2d 1403, 1413-16 (11th Cir. 1985); *Bester v. Tuscaloosa City Bd. of Educ.*, 722 F.2d 1514, 1517 (11th Cir. 1984); *Debra P. v. Turlington*, 474 F. Supp. 244, 249-57 (M.D. Fla. 1979) *aff'd in part and vacated in part*, 644 F.2d 397 (5th Cir. 1981).

APPENDIX B: Glossary of Test Measurement Terms

This glossary is provided as a plain language reference to assist readers in understanding commonly used test measurement terms used in this guide or terms relevant to issues discussed in the guide. For additional relevant information, readers are encouraged to review the Glossary in the *Joint Standards*, as well as the appropriate chapters in the *Joint Standards*.

Accommodation—A change in how a test is presented, in how a test is administered, or in how the test taker is allowed to respond. This term generally refers to changes that do not substantially alter what the test measures. The proper use of accommodations does not substantially change academic level or performance criteria. Appropriate accommodations are made in order to level the playing field, i.e., to provide equal opportunity to demonstrate knowledge.

Achievement level/ proficiency levels—Descriptions of a test taker's competency in a particular area of knowledge or skill, usually defined as ordered categories on a continuum, often labeled from "basic" to "advanced," that constitute broad ranges for classifying performance.

Alternate assessment—An assessment designed for those students with disabilities who are unable to participate in general large-scale assessments used by a school district or state, even when accommodations or modifications are provided. The alternate assessment provides a mechanism for students with even the most significant disabilities to be included in the assessment system.

Assessment—Any systematic method of obtaining information from tests or other sources, used to draw inferences about characteristics of people, objects, or programs.

Bias—In a statistical context, a systematic error in a test score. In discussing test fairness, bias may refer to construct underrepresentation or construct irrelevant components of test scores. Bias usually favors one group of test takers over another.

Bilingual—The characteristic of being relatively proficient in two languages.

Classification accuracy—The degree to which neither false positive nor false negative categorizations and diagnoses occurs when a test is used to classify an individual or event.

Composite score—A score that combines several scores according to a specified formula.

Content areas—Specified subjects in education, e.g. language arts, science, mathematics, or history.

Content domain—The set of behaviors, knowledge, skills, abilities, attitudes or other characteristics to be measured by a test, represented in a detailed specification, and often organized into categories by which items are classified.

Content standard—Statements which describe expectations for students in a subject matter at a particular grade or at the completion of a level of schooling.

Content validity—Validity evidence which analyzes the relationship between a test's content and the construct it is intended to measure. Evidence based on test content includes logical and empirical analyses of the relevance and representativeness of the test content to the defined domain of the test and the proposed interpretations of test scores.

Construct—The concept or the characteristic that a test is designed to measure.

Construct equivalence—1. The extent to which the construct measured by one test is essentially the same as the construct measured by another test. 2. The degree to which a construct measured by a test in one cultural or linguistic group is comparable to the construct measured by the same test in a different cultural or linguistic group.

Construct irrelevance—The extent to which test scores are influenced by factors that are irrelevant to the construct that the test is intended to measure. Such extraneous factors distort the meaning of test scores from what is implied in the proposed interpretation.

Constructed response item—An exercise for which examinees must create their own responses or products rather than choose a response from an enumerated set. Short-answer items require a few words or a number as an answer, whereas extended-response items require at least a few sentences.

Construct underrepresentation—The extent to which a test fails to capture important aspects of the construct that the test is intended to measure. In this situation, the meaning of test scores is narrower than the proposed interpretation implies.

Criterion validity—Validity evidence which analyzes the relationship of test scores to variables external to the test. External variables may include criteria that the test is expected to be associated with, as well as relationships to other tests hypothesized to measure the same constructs and tests measuring related constructs. Evidence based on relationships with other variables addresses questions about the degree to which these relationships are consistent with the construct underlying the proposed test interpretations. See predictive validity.

Criterion-referenced—Scores of students referenced to a criterion. For instance, a criterion may be specific, identified knowledge and skills that students are expected to master. Academic content standards in various subject areas are examples of this type of criterion.

Criterion-referenced test—A test that allows its users to make score interpretations in relation to a functional performance level, as distinguished from those interpretations that are made in relation to the performance of others. Examples of criterion-referenced interpretations include comparison to cutscores, interpretations based on expectancy tables, and domain-referenced score interpretations.

Cutscore—A specified point on a score scale, such that scores at or above that point are interpreted or acted upon differently from scores below that point. See *Performance standard*.

Discriminant validity—Validity evidence based on the relationship between test scores and measures of different constructs.

Error of measurement—The difference between an observed score and the corresponding true score or proficiency. This unintended variation in scores is assumed to be random and unpredictable and impacts the estimate of reliability of a test.

False negative—In classification, diagnosis, or selection, an error in which an individual is assessed or predicted not to meet the criteria for inclusion in a particular group but in truth does (or would) meet these criteria.

False positive—In classification, diagnosis, or selection, an error in which an individual is assessed or predicted to meet the criteria for inclusion in a particular group but in truth does not (or would not) meet these criteria.

Field test—A test administration used to check the adequacy of testing procedures, generally including test administration, test responding, test scoring, and test reporting. A field test is generally more extensive than a pilot test. See *Pilot test*.

High-stakes decision for students—A decision whose result has important consequences for students.

Internal consistency estimate of reliability—An index of the reliability of test scores derived from the statistical interrelationships of responses among item responses or scores on separate parts of a test.

Inter-rater agreement—The consistency with which two or more judges rate the work or performance of test takers; sometimes referred to as *inter-rater reliability*.

Local evidence—Evidence (usually related to reliability or validity) collected for a specific and particular set of test takers in a single institution, district, or state, or at a specific location.

Local norms—Norms by which test scores are referred to a specific, limited *reference population* of particular interest to the test user (e.g., institution, district, or state); local norms are not intended as representative of populations beyond that setting.

Norm-referenced—Scores of students compared to a specified reference population.

Norm-referenced test—A test that allows its users to make score interpretations of a test taker's performance in relation to the performance of other people in a specified reference population.

Norms—Statistics or tabular data that summarize the distribution of test performance for one or more specified groups, such as test takers of various ages or grades. The group of examinees represented by the norms is referred to as the reference population. Norm reference populations can be a local population of test takers, e.g. from a school, district or state, or it can represent a larger population, such as test takers from several states or throughout the country.

Percentile rank—Most commonly, the percentage of scores in a specified distribution that fall below the point at which a given score lies. Sometimes the percentage is defined to include scores that fall at the point; sometimes the percentage is defined to include half of the scores at the point.

Performance assessments—Product- and behavior-based measurements based on settings designed to emulate real-life contexts or conditions in which specific knowledge or skills are actually applied.

Performance standard—1. An objective definition of a certain level of performance in some domain in terms of a cut score or a range of scores on the score scale of a test measuring proficiency in that domain. 2. A statement or description of a set of operational tasks exemplifying a level of performance associated with a more general content standard; the statement may be used to guide judgements about the location of a cut score on a score scale. The term often implies a desired level of performance. *See Cutscores.*

Pilot test—A test administered to a representative sample of test takers to try out some aspects of the test or test items, such as instructions, time limits, item response formats, or item response options. *See Field test.*

Portfolio assessments—A systematic collection of educational or work products that have been compiled or accumulated over time, according to a specific set of principles.

Precision of measurement—A general term that refers to a measure's sensitivity to error of measurement.

Predictive validity—Validity evidence that analyzes the relationship of test scores to variables external to the test that the test is expected to predict. Predictive evidence indicates how accurately test data can predict criterion scores that are obtained or outcomes that occur at a later time. When test scores are used to predict a dichotomous criterion, such as a diagnosis, false positive and false negative errors can occur. *See Criterion evidence of validity; False positive error; False negative error.*

Random error—An unsystematic error; a quantity (often observed indirectly) that appears to have no relationship to any other variable.

Reference population—The population of test takers represented by test norms. The sample on which the test norms are based must permit accurate estimation of the test score distribution for the reference population. The reference population may be defined in terms of size of the population (local or larger), examinee age, grade, or clinical status at time of testing, or other characteristics.

Reliability—The degree to which test scores for a group of test takers are consistent over repeated applications of a measurement procedure and hence are inferred to be dependable, and repeatable for an individual test taker; the degree to which scores are free of errors of measurement for a given group.

Sample—A selection of a specified number of entities called **sampling units** (test takers, items, schools, etc.) from a large specified set of possible entities, called the **population**. A **random sample** is a selection according to a random process, with the selection of each entity in no way dependent on the selection of other entities. A **stratified random sample** is a set of random samples, each of a specified size, from several different sets, which are viewed as **strata** of the population.

Sampling from a domain—The process of selecting test items to represent a specified universe of performance.

Score—Any specific number resulting from the assessment of an individual; a generic term applied for convenience to such diverse measures as test scores, absence records, course grades, ratings, and so forth.

Scoring rubric—The established criteria, including rules, principles, and illustrations, used in scoring responses to individual items and clusters of items. The term usually refers to the scoring procedures for assessment tasks that do not provide enumerated responses from which test takers make a choice. Scoring rubrics vary in the degree of judgement entailed, in the number of distinct score levels defined, in the latitude given scorers for assigning intermediate or fractional score values, and in other ways.

Selection—A purpose for testing that results in the acceptance or rejection of applicants for a particular educational opportunity.

Sole criterion—When only one standard (such as a test score) is used to make a judgement or a decision. This can include a step-wise decision-making procedure where students must reach or exceed one criterion (such as a cutscore of a test) independent of or before other criteria can be considered.

Speed test—A test in which performance is measured primarily or exclusively by the time to perform a specified task, or the number of tasks performed in a given time, such as tests of typing speed and reading speed.

Standards-based assessment—Assessments intended to represent systematically described content and performance standards.

Systematic error—A score component (often observed indirectly), not related to the test performance, that appears to be related to some salient variable or sub-grouping of cases in empirical analyses. This type of error tends to increase or decrease observed scores consistently in members of the subgroup or levels of the salient variable. See *Bias*.

Technical manual—A publication prepared by test authors and publishers to provide technical and psychometric information on a test.

Test—An evaluative device or procedure in which a sample of an examinee's behavior in a specified domain is obtained and subsequently evaluated and scored using a standardized process.

Test developer—The person(s) or agency responsible for the construction of a test and for the documentation regarding its technical quality for an intended purpose.

Test development—The process through which a test is planned, constructed, evaluated and modified, including consideration of content, format, administration, scoring, item properties, scaling, and technical quality for its intended purpose.

Test documents—Publications such as test manuals, technical manuals, user's guides, specimen sets, and directions for test administrators and scorers that provide information for evaluating the appropriateness and technical adequacy of a test for its intended purpose.

Test manual—A publication prepared by test developers and publishers to provide information on test administration, scoring, and interpretation and to provide technical data on test characteristics.

Validation—The process through which the validity of the proposed interpretation of test scores is evaluated.

Validity—The degree to which accumulated evidence and theory support specific interpretations of test scores entailed by proposed uses of a test.

Validity argument—An explicit scientific justification of the degree to which accumulated evidence and theory supports the proposed interpretation(s) of test scores.

Validity evidence—Systematic documentation that empirically or theoretically demonstrates, under the specific conditions of the individual analysis, to which extent, for whom, and in which situations test score inferences are valid. No single piece of evidence is sufficient to document validity of test scores; rather, aspects of validity evidence must be accumulated to support specific interpretations of scores.

Validity evidence for relevant subgroups—Validity results disaggregated by subgroups, e.g. by race/ethnicity, or by disability or LEP status. This type of evidence is appropriate generally when credible research suggests that interpretations of the test scores may differ by subgroup. For instance, if a test will be used to predict future performance, validity evidence should document that the scores are as valid a predictor of the intended performance for one subgroup as for another.

APPENDIX C: Accommodations Used by States

Table 1
Accommodations for Limited English Proficient Students

PRESENTATION FORMAT

Translation of directions into native language
Translation of test into native language
Bilingual version of test (English and native language)
Further explanation of directions
Plain language editing
Use of word lists/ dictionaries
Bilingual dictionary
Large print

ADMINISTRATION FORMAT

Oral reading in English
Oral reading in native language
Person familiar to students administers test
Clarification of directions
Use of technology
Alone, in study carrel
Separate room
With small group
Extended testing time
More breaks
Extending sessions over multiple days

RESPONSE FORMAT

Allow student to respond in writing in native language
Allow student to orally respond in native language
Allow student to orally respond in English
Use of technology

OTHER

Out-of-level testing
Alternate scoring of writing test

Adapted from: Council of Chief State School Officers, *Annual Survey: State Student Assessment Programs*, Washington D.C., 1999

Table 2
Accommodations for Students with Disabilities

PRESENTATION FORMAT

Braille edition
Large-print editions
Templates to reduce visual field
Short-segment testing booklets
Key words highlighted in directions
Reordering of items
Use of spell checker
Use of word lists/dictionaries
Translated into sign language

ADMINISTRATION FORMAT

Oral reading of questions
Use of magnifying glass
Explanation of directions
Audiotape directions or test items
Repeating of directions
Interpretation of directions
Videotape in American Sign Language
Interpreter signs test in front of classroom/student
Signing of directions
Amplification equipment
Enhanced lighting
Special acoustics
Alone in study carrel
Individual administration
In small groups
At home with appropriate supervision
In special education classes separate room
Off campus
Interpreter with teacher facing student; student in front of classroom
Adaptive furniture
Use place marker
Hearing aids
Student wears noise buffers
Administrator faces student
Specialized table

Auditory trainers
Read questions aloud to self
Colored transparency
Assist student in tracking by placing students finger on item
Typewriter device to screen out sounds
Extended testing time
More breaks
Extending sessions over multiple days
Altered time of day that test is administered

RESPONSE FORMAT

Mark responses in booklet
Use template for recording
Point to response
Lined paper
Use sign language
Use typewriter/computer/ word processor
Use Braille writer
Oral response, use of scribe
Alternative response methods, use of scribe
Answers recorded on audiotape
Administrator checks to ensure that student is placing responses in correct area
Lined paper for large script printing
Communication board

OTHER

Out-of-level testing

Adapted from: Council of Chief State School Officers, *Annual Survey: State Student Assessment Programs*, Washington D.C., 1999

APPENDIX D: Compendium of Federal Statutes and Regulations

This compendium provides a description of the federal nondiscrimination statutes and regulations that are relevant to testing issues and constitute the primary sources of legal authority in the guide. Specifically, this appendix primarily provides information on federal civil rights laws, including Title VI, Title IX, Section 504, and Title II of the Americans with Disabilities Act.

A. Title VI

Title VI of the Civil Rights Act of 1964, 42 U.S.C. § 2000d, prohibits race and national origin discrimination in programs and activities that receive federal financial assistance. For the regulations issued by the Department of Education implementing Title VI, see 34 C.F.R. Part 100. Under the Civil Rights Restoration Act of 1987, OCR has institutionwide jurisdiction over the recipient of federal funds. See 42 U.S.C. § 2000d-4 (1989).

The Title VI statute bars only intentionally discriminatory conduct. However, the regulations promulgated under Title VI prohibit the use of neutral criteria having disparate effects unless the criteria are educationally justified and there are no alternative practices available that are equally effective in serving the institution's goals and result in less disparate effects. See *Guardians Ass'n v. Civil Service Comm'n*, 463 U.S. 582 (1983).

The regulations implementing Title VI do not specifically address the use of tests and assessment procedures, but bar discrimination based on race, color or national origin in any service, financial aid or other benefit provided by the recipient. 34 C.F.R. § 104.3(b)(2), which prohibits criteria or methods of administration having an unjustified discriminatory effect, is often applied in testing cases.

See also 34 C.F.R. § 100, Appendix B, part K (Guidelines for Eliminating Discrimination and Denial of Services on the Basis of Race, Color, National Origin, Sex, and Handicap in Vocational Education Programs) ("if a recipient can demonstrate that criteria [that disproportionately exclude persons of a particular race, color, national origin, sex, or disability] have been validated as essential to participation in a given program and that alternative equally valid criteria that do not have such a disproportionate adverse effect are unavailable, the criteria will be judged nondiscriminatory. Examples of admission criteria that must meet this test or assessment procedure are . . . interest inventories . . . and standardized test or assessment procedures").

B. Title IX

Title IX of the Education Amendments of 1972, 20 U.S.C. §§ 1681 *et seq.*, prohibits sex discrimination in education programs that receive federal financial assistance. For the regulations issued by the Department of Education implementing Title IX, see 34 C.F.R. Part 106. As under Title VI, OCR, per the Civil Rights Restoration Act of 1987, generally has institutionwide jurisdiction over the recipient of federal funds. See 42 U.S.C. § 2000d(4) (1989).

In addition to general prohibitions against discrimination, the regulations implementing Title IX specifically prohibit the discriminatory use of test or assessment procedures in admissions, 34 C.F.R. § 106.21, employment, 34 C.F.R. § 106.52, and counseling 34 C.F.R. § 106.36.

See also 34 C.F.R. § 100, Appendix B, part K (Guidelines for Eliminating Discrimination and Denial of Services on the Basis of Race, Color, National Origin, Sex, and Handicap in Vocational Education Programs), discussed above in relation to Title VI.

C. Section 504 of the Rehabilitation Act of 1973

Section 504 prohibits discrimination based on disability in programs and activities receiving federal financial assistance. OCR enforces Section 504 and its regulations in education programs. The regulations implementing Section 504 contain certain sections that are particularly relevant to testing situations:

34 C.F.R. § 104.4(b)(4) prohibits criteria or methods of administration that have the effect of discriminating against qualified persons with disabilities.

34 C.F.R. § 104.42(b)(2) prohibits admissions procedures by higher educational institutions that make use of any test or criterion for admission that has a disproportionate, adverse impact on qualified individuals with disabilities unless (1) the test or criterion, as used by the institution, has been validated as a predictor of success in the education program or activity and (2) alternate tests or criteria that have a less disproportionate, adverse impact are not shown to be available. 34 C.F.R. § 104.42(b)(3) requires admissions tests used by post-secondary institutions to be selected and administered so as best to ensure that, when a test is administered to an applicant with a disability, the test results accurately reflect the applicant's aptitude or achievement, rather than reflecting the student's disability (except where disability-related skills are the factors the test purports to measure). 34 C.F.R. §§ 104.44(a) and 104.44(d) require higher education institutions to provide adjustments or accommodations and auxiliary aids and services that enable the student to demonstrate the knowledge and skills being tested.

34 C.F.R. § 104.44(a) states that academic requirements that the institution can demonstrate are essential to the program of instruction or to any directly related licensing requirement will not be regarded as discriminatory.

34 C.F.R. § 104.35(b) requires public elementary and secondary education programs to individually evaluate a student before classifying the student as having a disability or placing the student in a special education program; tests used for this purpose must be selected and administered so as best to ensure that the test results accurately reflect the student's aptitude or achievement or other factor being measured rather than reflecting the student's disability, except where those are the factors being measured. These provisions also require that tests and other evaluation materials include those tailored to evaluate the specific areas of educational need and not merely those designed to provide a single intelligence quotient.

D. Title II of the Americans with Disabilities Act (ADA)

Title II of the Americans with Disabilities Act of 1990 (ADA), 42 U.S.C. § 12134, prohibits discrimination on the basis of disability by public entities. Regulations implementing Title II, issued by the U.S. Department of Justice, can be found at 28 C.F.R. Part 35. OCR enforces Title II as to public schools and colleges. Like the Section 504 regulations, the regulations implementing Title II prohibit "criteria and methods of administration which have the effect of discriminating" against qualified persons with disabilities. See 28 C.F.R. § 35.130(b)(3). The regulations also require public entities to make reasonable accommodations to policies, procedures, and practices when the modifications are necessary to avoid discrimination unless the public entity can demonstrate that the modification would fundamentally alter the nature of the service, program, or activity. See 28 C.F.R. § 35.130(b)(7).

E. Individuals with Disabilities Education Act (IDEA)

Although not a discrimination law per se, IDEA contains important provisions related to testing students with disabilities in elementary and secondary schools. IDEA is enforced by the Office of Special Education Programs in the U.S. Department of Education. As amended in 1997, IDEA requires inclusion of students with disabilities in state and districtwide assessment programs, with appropriate accommodations, if necessary, unless the student's individual education team decides that participation in all or part of the testing program is not appropriate. The student's individualized education program (IEP) should also state 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. 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 also requires state or local education agencies

to develop guidelines for the alternate assessment of the relatively small number of students with disabilities 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. See 20 U.S.C. §§ 1412(a)(16) and (17), 1413(a)(6), and 1414(d)(1)(A) and (d)(6)(A)(ii); and regulations at 34 C.F.R. §§ 300.138, 300.139, 300.240, and 300.347.

APPENDIX E: Resources and References

Office for Civil Rights U.S. Department of Education

Minority Students and Special Education: Legal Approaches for Investigation (1995).
Provides an overview of the legal theories and approaches employed in OCR investigations examining disproportionate representation of minority students in special education.

Policy Update on Schools' Obligations Toward National-Origin-Minority Students With Limited-English Proficiency (1991).

Used by OCR staff to determine schools' compliance with their Title VI obligation to provide any alternative language programs necessary to ensure that national-origin-minority students with limited English proficiency have meaningful access to programs. Provides additional guidance for the December 1985 and May 1970 memoranda.

The Office for Civil Rights' Title VI Language-Minority Compliance Procedures (1985).
Focuses on the treatment of limited English proficient students in programs that received funds from the Department.

Identification of Discrimination and Denial of Services on the Basis of National Origin (May, 1970) 35 Fed. Reg. 11595.

Clarifies school district responsibilities to limited English proficient students. Memo was the foundation for the U.S. Supreme Court decision *Lau v. Nichols* and was affirmed in that decision.

Office of Elementary and Secondary Education U.S. Department of Education

Peer Reviewer Guidance for Evaluating Evidence of Final Assessments Under Title 1 of the Elementary and Secondary Education Act (ESEA) (1999).

Informs the states about types of evidence that would be useful in determining the evaluation of assessments under Title 1.

Taking Responsibility for Ending Social Promotion (1999).

Provides strategies for preventing academic failure and gives information about how these strategies can be sustained through ongoing support for improvement.

Handbook for the Development of Performance Standards: Meeting the Requirements of Title 1 (with Council of Chief State School Officers) (1998).

Describes the best practices and current research on the development of academic performance standards for K-12.

Standards, Assessments and Accountability (1997).

Overview of the major provisions under Title 1 of the Elementary and Secondary Education Act.

**National Research Council
National Academy Press, Washington D.C.**

High Stakes: Testing for Tracking, Promotion and Graduation (Jay P. Heubert & Robert M. Hauser eds., 1999).

Discusses how tests should be planned, designed, implemented, reported and used for a variety of educational policy goals. Focuses on the uses of tests that make high-stake decisions about individuals and on how to ensure appropriate test use.

Myths and Tradeoffs: The Role of Tests in Undergraduate Admissions (Alexandra Beatty, M.R.C. Greenwood & Robert L. Linn eds., 1999).

Four recommendations regarding test use for admission are made to colleges and universities, including a warning to schools to avoid using scores as more precise and accurate measures of college readiness than they are. One recommendation is made to test producers, which is to make clear the limitations of the information that the scores provide.

Testing, Teaching and Learning: A Guide for States and School Districts (Richard F. Elmore & Robert Rothman eds., 1999).

Practical guide to assist states and school districts in developing challenging standards for student performance and assessment as specified by Title I. Discusses standards-based reform and specifies components of an education improvement system, which are standards, assessments, accountability and monitoring the conditions of instruction.

Improving America's Schooling for Language Minority Children: A Research Agenda (Diane August & Kenji Hakuta eds., 1997).

Summarization of extensive study of limited English proficient students. Gives state of knowledge review and identifies research agenda for future study. Includes discussion of student assessment and program evaluation.

Educating One and All: Students with Disabilities and Standards-Based Reform

(Lorraine M. McDonnell, Margaret J. McLaughlin & Patricia Morison eds., 1997). Twelve recommendations are given regarding how to integrate students with disabilities in standards-based reform, including: participation of students with disabilities should be maximized; that any test alterations must be individualized and have a compelling educational justification; include these students' test results in any accountability system; ensure opportunity for students with disabilities to learn the material tested; and use the IEP process for decision-making on the participation of individual students. Recommendations for policy-makers include: revising policies that discourage the inclusion of students with disabilities in high-stake tests; giving parents enough information to make informed choices about participation; monitoring possible unanticipated consequences of participation, both for standardized testing and for students with disabilities; designing realistic standards; and designing a long-term research agenda.

The Use of I.Q. Tests in Special Education Decision-Making and Planning: Summary of Two Workshops (Patricia Morison, S.H. White & Michael J. Feuer eds., 1996). Report provides a synthesis of the key themes and ideas discussed at workshops, including: an overview of legal, policy and measurement issues in use of IQ tests in special education; validity and fairness of IQ testing for student classification and placement; alternative assessment methods used in combination with or as substitutes for IQ tests.

Responsible Test Use: Case Studies for Assessing Human Behavior (Lorraine D. Hyde, Gary J. Robertson & Samuel E. Krug, et. al., eds., 1993). Casebook for professionals using educational and psychological test data, which was developed to apply principles to proper test interpretation and actual test use. Cases are organized under eight sections: general training, professional responsibility training, test selection, test administration, test scoring and norms, test interpretation, reporting to clients and administrative or organization policy issues.

Test Measurement Standards

American Educational Research Association, American Psychological Association & National Council on Measurement in Education, *Standards of Educational and Psychological Testing* (1999).

Provides criteria for the evaluation of tests, testing practices, and the effects of test use. Begins with discussion of the test development process, which focuses on test developers, and moves to specific test uses and applications, which focus on test users. One chapter centers on test takers.

National Council on Measurement in Education, *Code of Professional Responsibilities in Educational Measurement* (1995).

Association for Measurement and Evaluation in Counseling and Development, *Responsibilities of Users of Standardized Tests* (1992).

Joint Committee on Testing Practices, *Code of Fair Testing Practices in Education* (1988).

Measurement Texts

Educational Measurement (Robert L. Linn, ed., 3rd ed. 1989).

Includes 11 chapters, including Messick's classic chapter on validity, and organizes them in two parts: theory and general principles; and construction, administration and scoring.

Samuel Messick, *Validity of Psychological Assessment: Validation of Inferences from Persons' Responses and Performances as Scientific Inquiry into Score Meaning*, *American Psychologist* 50(9) (September 1995).

Gives a new cohesive definition of validity that looks at score meaning and social values. Six perspectives of construct validity are defined: content, substantive, structural, generalizability, external and consequential.

Martha Thurlow, Judy Elliott & Jim Ysseldyke, *Testing Students With Disabilities* (1998).

This document provides guidance about how students with disabilities should be included in large-scale tests, considerations about how to select the appropriate accommodations for which students, and discussions about the role of state and local educators in ensuring proper test use, the use of alternate tests, and appropriate reporting considerations.

Rebecca J. Kopriva, Council of Chief State School Officers, *Ensuring Accuracy in Testing for English Language Learners* (2000).

This resource provides guidance to states, districts, and test publishers about developing, selecting, or adapting large-scale, standardized assessments of educational achievement that are appropriate and valid for English language learners. The guide's practical recommendations identify the "who, what, when, why and how" associated with developing, selecting, or adapting tests for institution use, including how to select the appropriate accommodations for which students, how to collect appropriate validity evidence, and a discussion of salient reporting considerations.

Test Publisher Materials

Most test publishers produce materials that explain the appropriate use of their tests. We encourage interested readers to obtain these materials from the publishers of the tests they administer or from publishers of tests in which they are interested. Readers can also contact the Association of Test Publishers, 655 15th St. NW, Washington, D.C., 20005, telephone 202-857-8444 for more information.

Other Resources

There are many books and other materials that might be helpful to educators and policymakers as they develop policies, and design and implement programs which include the use of tests in making high-stakes decisions for students. The following web sites will provide additional information and links to some of these resources.

Council for Chief State School Officers

<http://www.CCSSO.org>

The National Center on Education Outcomes

<http://www.coled.umn.edu/NCEO>

Center for Evaluation, Research, Standards and Student Testing

<http://cresst96.cse.ucla.edu>

National Clearinghouse for Bilingual Education

<http://www.ncbe.gwu.edu>