

Educ -
Civ. Rts.

A Rocky Path for Russia

THE CHALLENGE of a few thousand Muslim separatist guerrillas in the southern Russian province of Dagestan once again raises the question of whether Russia can hold together. The empire controlled from Moscow has been shrinking for a decade. With the collapse of the Berlin Wall in 1989, the captive nations of Central and Eastern Europe slipped their bonds. With the collapse of the Soviet Union in 1991, Russia suddenly had 14 new neighbors and far less territory. Moscow relinquished effective control of its southern province of Chechnya after losing a war with that breakaway republic in 1994 and 1995. Has Russia (still the world's-largest country) now shrunk to its "natural" borders or is further unraveling in the cards?

Elements of the latest challenge must be familiar to the Moscow power brokers who failed to prevent Chechnya's secession. On one side are highly motivated insurgents fighting in familiar terrain; on the other, poorly armed, poorly fed Russian draftees transported to treacherous mountains that they have no interest in defending or recapturing. Chechen guerrillas often shot Russian troops by night with weapons that Russian troops had sold them by day. This time around the Russian armed forces are no less corrupt, and certainly no less impoverished. As during the war in Chechnya, Russian leaders proclaim imminent victory in

plain contradiction of the evident facts.

Both Dagestan and Chechnya are predominantly Muslim entities in the Caucasus region. But the ethnic Chechen population was fairly united in its desire for independence, and Dagestan is a patchwork of ethnic and tribal affiliation that is not united in anything. Many people there fear that a break from Russia would lead to internal conflicts. Much of Dagestan's elite has traditionally aligned itself with Moscow.

But—and here we return to the central question of Russia's viability—that loyalty in the past has been bought, at least in part. Moscow took money from the few provinces that operate profitably—primarily oil- and diamond-producing areas—and recycled it to provinces such as Dagestan that never paid their own way. Now the capital has far less power to extract taxes from those who can pay and so fewer inducements to bind those who cannot.

The answer to the central question, in other words, does not lie in Dagestan's mountains as much as in Moscow itself. If Russia can put its economic reform on track and protect its fragile democratic institutions, most Russians will want to remain just that—Russian. If the economy spirals downward and corruption becomes a permanent fixture, Dagestan may seem a few years from now to have been nothing but a harbinger.

If Not Tests?

THE DRUMBEAT on raising educational standards has been one of this administration's better contributions to public policy. In recent years, more school districts have bucked inertia and moved toward addressing persistent achievement gaps between white and minority students and between rich and poor. They have adopted tough tests that, because they have real consequences for students' prospects, force underperforming students and teachers alike to make extraordinary efforts to do better. Yet earlier this summer, the Department of Education took a step that could undercut that useful pressure.

The department's Civil Rights Office circulated draft legal guidelines on "nondiscrimination in high-stakes testing." The draft warned schools that educational testing of this kind may constitute a civil rights violation if it denies a benefit (such as promotion or graduation) disproportionately to female or minority students. A test that exposes an achievement gap between black and white students, in other words, could be open to legal challenge on that basis alone.

The appearance of the draft caused a predictable outcry. Many feared it could speed challenges to admissions tests such as the SAT, encourage lawsuits and even discourage schools from using standardized testing. The department protested that the document—now undergoing revision—breaks no new legal ground and was intended merely to help schools keep themselves out of a danger zone as they adopt tests with real consequences. A school or

system may go on using a standardized test that shows lower performance on average by girls or minorities—the draft paper reassures—if the school can show the test is "educationally valid," and if an outside party cannot prove that another type of measurement or test would meet the educational purposes of the first while having less of a "disparate impact."

This may not be so much reassuring as intimidating to a school contemplating the prospect of a challenge. A revised guide could, and should, offer much more on what testing practices are *not* liable to such challenges. But the draft's approach also misses the extent to which the wider battle over race and education has moved past the question of whether achievement gaps are an artifact of racist testing. The forced dismantling of college affirmative action plans in California and Texas—plans that sought to modify just such "disparate impact" by using factors other than grades and test scores in college admissions—has focused attention on the problem that produced the disparities—that of unequal K-12 preparation. It can be fixed only by long-running efforts to make poor schools more rigorous.

The civil rights issue here is the need to ensure that minority kids in poor schools get the help they need to pass the tests. They need an equal chance to learn and be taught the material, time to prepare for "high-stakes" tests and, in some places, the chance to take them more than once. What they don't need is quicker recourse to legal action that takes away the pressure to improve.

Managed Care Maneuvering

THE HIGH-WIRE act occasioned by the lack of a clear majority in the House—the odd and disabling fact that the majority leadership spends much of its time defending minority positions against the House's will—continues. The issue now is managed care. Speaker Dennis Hastert's margin is such that, if no Democrats come his way, he can afford to lose only four Republicans and still maintain control. On managed care, he is threatened with the loss of at least 21 who have signed on as supporters of a regulatory effort he seeks to deflect. He is looking for a way to bring them back. He wants to hold a vote, lest Democrats charge him and the party with keeping a popular—not to mention mostly worthy—bill from becoming law. But a loss if he lets the bill onto the floor would embarrass him and disappoint important Republican constituencies alike.

It's an awkward choice and not the only issue on which Mr. Hastert now finds himself thus pinched. He is bringing a campaign finance reform bill to the floor next month because members of his own party threatened to join Democrats in signing a discharge petition that would have forced it to the floor had he refused. The leadership will try to keep it from passing by trumping it with an innocuous bill. If that fails, opponents will try poison-pill amendments, whose adoption would likely cause Democrats to vote against the underlying measure; then Republicans might not have to bear the blame for its defeat.

The speaker has had to resort to similar tactics—appearing to push in one direction while pushing in another—to pass the tax and spending bills that are supposed to summarize what Republicans stand for, and thereby what is at stake, as the country makes the turn into next year's elections. To finance the tax cut it wants to grant, the leadership is supporting spending cuts for which it lacks the votes. The speaker was able to put together a majority for the tax

cut only because of the understanding on all sides that the president instantly would veto it as excessive. Moderate Republicans opposed to the bill were able to vote aye, secure in the knowledge that it would be only a gesture. Likewise on appropriations, where for lack of the necessary votes, Mr. Hastert basically has deferred the cuts that the financing of the tax bill would require. The hope is to work out a deal with the president later in the year in which he will bear the responsibility for what the Republicans will denounce as spending increases but nonetheless support.

The managed care issue is particularly difficult for Republicans, in that it is forcing them to choose between two traditional constituencies, physicians and employers. The docs—not all, but many—detest managed care for undercutting their prerogatives and income. They'd be happy to see the industry regulated, as Democrats seek to do in the name of patients' rights. But employers who pay for employee health insurance support aggressive managed care because it cuts their costs. The insurance industry, which has morphed into the managed care industry, also opposes serious regulation.

The Republican leadership likewise favors light—some would say, token—regulation; the Senate already has passed a token bill over Democratic opposition. But a handful of House Republicans who are members of the medical profession have led a mini-revolt to the other side. A compromise has been worked out by Reps. Charles Norwood, a Republican dentist, and John Dingell, ranking Democrat on the Commerce Committee, that in some respects is an improvement on the original Democratic bill; it's amazing what good ideas old-fashioned legislative give-and-take sometimes can produce when the members try. The administration supports the compromise. That's what Mr. Hastert still hopes to defeat. It's as good an emblem as any for the year thus far.

The Washington Post

MONDAY, AUGUST 23, 1999

*Edue -
Civil Rights*

*Bruce -
This doesn't say much. It just
The legal framework and principles
much like a treatise would. It
doesn't commit us to any
JUL 28 1997*

TO : Senior Officers
U.S. Department of Education

CC : Jessica Levin
Steve Winnick
Susan Craig
Joe Conaty

*particular embrocement approach
in strategy. i.e. - don't worry.
Elena*

FROM : Norma V. Cantú
Assistant Secretary
for Civil Rights

Norma V. Cantú

Arthur L. Coleman
Deputy Assistant Secretary
for Civil Rights

Arthur L. Coleman

SUBJECT : OCR's Investigative Guidance on Fairness in Testing/Deadline of August 15

Attached for your review is a final draft of OCR's Investigative Guidance on Fairness in Testing. The Guidance is designed to provide OCR's attorneys and investigators with an explanation of the legal and conceptual framework needed for understanding the issues raised by challenges under Title VI of the Civil Rights Act of 1964 and Title IX of the Educational Amendments of 1972 to testing and assessment practices. (Notably, OCR has applied the principles set forth in this Guidance in a manner fully consistent with the Department of Education's emphasis on high standards.) It also may be used as a point of reference by others who seek to understand the principles that guide any determination about uses of particular tests under Title VI and Title IX.

Since your previous review of the draft Guidance many months ago, the Guidance has been reviewed by the Board on Testing and Assessment (BOTA) of the National Academy of Sciences. The Guidance has been revised in response to comments from BOTA and in response to additional refining comments from OCR, the Office of the General Counsel, the Office of the Deputy Secretary, and the Office of Educational Research and Improvement. We are grateful for the collaboration that has led to this statement of principles that govern our work.

The substantive changes that have been made to the Guidance since its earlier circulation are technical in nature. These changes do not alter the basic principles or standards about which we achieved agreement many months ago; the changes are, for the most part, minor. For example, there is further explanation of concepts related to test validation and reliability.

Specifically, a definition of construct validity has been added and the discussion of reliability is now in a separate section. Two paragraphs have been added to the discussion regarding cutoff scores. Similarly minor changes have been made to the discussion regarding statistical analyses necessary to establish disparate impact. Also, a glossary on concepts fundamental to test validation has been added (Tab C).

Thank you for your review of the Guidance. Because we would like to issue the Guidance in final in late August, we would appreciate receiving comments by August 15, 1997. Please send your comments to Jeanette Lim of OCR. Ms. Lim can be contacted by e-mail, by telephone at 205-8635, or by fax at 260-3040.

Attachment

TO : All OCR Staff
FROM : Norma V. Cantú
Assistant Secretary for Civil Rights
SUBJECT : Fairness in Testing

FAIRNESS IN TESTING: AN OVERVIEW

I disagree with the proposition that there are inherent racially based differences in the capacity of the American people to reach their full potential.

President Bill Clinton, October 21, 1994

An invalid test cannot measure merit.

Walls v. Mississippi State Dept. of Public Welfare, 542 F. Supp. 281, 311 (N.D. Miss. 1982), aff'd in relevant part, 730 F. 2d 306 (1984).

BACKGROUND

The issue of fairness in testing and other assessment practices is, at its core, a critical access to education issue. It was established as an OCR strategic plan high priority in 1993 following consultation both within and outside the Department. OCR investigators have encountered testing and assessment issues when they have addressed complaints on the subject of: the disproportionate representation of minorities in special education, the access of limited-English-proficient (LEP) children to equal educational opportunities, and desegregation, to name a few. This agency in the past has been guided on this issue by court decisions and principles established through OCR case-specific findings. There is no single document that has synthesized these decisions and findings.

OCR has developed this guidance in order to provide our attorneys and investigators with practical guidance on testing and assessment principles that lie at the core of Title VI of the Civil Rights Act of 1964 (Title VI) and Title IX of the Education Amendments of 1972 (Title IX) case law. This guidance also can be helpful in explaining to recipients the standards by which their compliance with Title VI and Title IX may be evaluated. This will better encourage voluntary compliance as well as a greater understanding of the testing and assessment parameters that guide OCR investigations.

Page 2 - Overview

The centrality of the issue of fairness in testing and assessment is properly viewed as consistent with, rather than contrary to, the push for high standards -- the cornerstone of many of the Department's initiatives. OCR's focus on fairness in testing and assessment helps to challenge the assumption that minority students perform poorly because they lack the ability or motivation to reach their full potential. All students need an educational system which both expects high performance and offers real and meaningful educational opportunities. Furthermore, the fact that a recipient's use of a test is violative of Title VI or Title IX does not mean that the automatic response to remedy the problem is to eliminate the test. Depending on the facts of a given case, there are many permissible responses to correcting a violation, which include: supplementing the use of the test with other assessment measures; revising the test instrument within a reasonable period of time to address compliance concerns; substituting the test with another available instrument that more appropriately measures what is intended to be measured; increasing remedial support; and enhancing learning opportunities for students to perform well on the test.

Significantly, OCR's focus on fairness in testing and assessment does not purport to mandate in any way the content of school curriculum. OCR's authority under Title VI and Title IX does not involve setting curricula or levels of instruction for schools. OCR's role, when necessary, is to ensure that students of all races and both sexes have equal access to the curricula and quality teaching that is offered to all.

It also is not OCR's intent or mandate to ensure equal outcomes by race, national origin, or gender. Rather, the focus of OCR, as reflected in the attached Guidance developed for OCR attorneys and investigators, is on principles of fairness in testing and assessment that are designed to ensure that all students have equal educational opportunities. Fairness in testing and assessment is essential to assuring that equal opportunities to educational excellence are provided regardless of race, national origin or gender so that all students may attain high standards.

SCOPE OF THE GUIDANCE

The attached Guidance provides an overview of the use, and misuse, of tests for making high stakes educational decisions, such as those that involve: school admissions; scholarship awards; evaluation for placement in gifted and talented programs, programs for LEP students; special education programs; vocational education counseling; and diploma awards. A glossary on a number of concepts fundamental to test validation is found at Tab C. In addition, more specific guidance is available on the placement of

Page 3 - Overview

minority students in special education. See Memorandum to All OCR Staff (July 6, 1995) (addressing Minority Students and Special Education).

The Guidance applies to norm-referenced and criterion-referenced tests¹ as well as professionally designed alternative forms of assessment, all of which are used for making "high stakes" educational decisions. The Guidance does not cover teacher-created classroom tests nor does it apply to modifications of tests and/or testing conditions required for the purpose of accommodating students with disabilities under Section 504 of the Rehabilitation Act or the Americans with Disabilities Act.

BASES OF THE GUIDANCE

Professional Standards

As reflected in the Guidance, OCR adheres to generally accepted professional standards for evaluating standardized tests, such as those described in the Standards for Educational and Psychological Tests prepared by a joint committee of the American Psychological Association, the American Educational Research Association, and the National Council on Measurement in Education; the Code of Fair Testing Practices in Education prepared by the Joint Committee on Testing Practices; and the EEOC Guidelines on Employee Selection Procedures. At OCR's request, the Board on Testing and Assessment of the National Academy of Sciences reviewed an earlier draft of the Guidance and provided comments in writing and during several discussions. Many of the comments have been incorporated into the Guidance, including further clarification of the scope of the Guidance, explanation of concepts relating to test validation and reliability, and the discussion of cutoff scores. Their comments have helped to ensure that the Guidance is consistent with professional standards and educationally sound.

¹ Norm-referenced tests are tests used to identify an individual's performance in relation to the performance of other people in a specified group on the same test. American Psychological Association Standards for Educational and Psychological Testing (1985) (APA Standards) at p. 92. Criterion-referenced tests allow users to make score interpretations in relation to a functional performance level. APA Standards at p. 90. In other words, criterion-referenced tests are designed to measure to what degree a learner has mastered a certain skill.

Page 4 - Overview

Legal Standards

The Guidance provides that when OCR investigates allegations involving discriminatory test use by a recipient, OCR may look at evidence under two separate legal theories of discrimination: disparate treatment and disparate impact. Each theory is based on settled legal principles and neither breaks any new legal ground.

A disparate treatment analysis is used by OCR to determine whether a policy or practice regarding testing is being applied differently by a recipient to an individual student or group of students of a particular race, national origin, or gender, without a legitimate, nondiscriminatory reason. Under this analysis OCR would determine, for example, whether black students and white students are being tested under different conditions or whether students with the same test scores are being treated differently by a recipient. If this is established, a recipient would have the opportunity to provide a legitimate, nondiscriminatory reason for the difference in treatment. If OCR determines that the reason is not a pretext for discrimination, the difference in treatment would be permissible under Title VI and Title IX.

Under a disparate impact analysis, OCR's focus would not be on the treatment of students, but rather on the effects of the application of a facially neutral policy regardless of whether the adverse consequences for a particular race, national origin, or gender were intended. Under a disparate impact analysis, further investigation will be triggered when the use of a test creates a significant difference in the granting or denial of benefits or opportunities on the basis of race, national origin or sex. Tests that have a disparate impact on the basis of race, national origin, or gender must be educationally necessary; otherwise, they are not permissible under Title VI or Title IX. Educational necessity requires a showing that a test is valid and reliable for the purpose for which it is being used and that there is no alternative available with less disparate impact that still serves the recipient's educational purpose. If validity and reliability have been demonstrated by the recipient, OCR has the burden of showing that an alternative is available that has less disparate impact and that would serve the recipient's educational purpose. The recipient's failure to use such an alternative would violate the civil rights statute(s) and should be remedied.

QUESTIONS NOT ADDRESSED BY THE GUIDANCE

What questions are not answered by this guidance? First and foremost, the investigative guidance, like other guidance issued by OCR, see 59 Fed. Reg. 11448 (March 10, 1993) (addressing Racial Incidents and Harassment), does not purport to define the

Page 5 - Overview

kinds of cases that may be targeted for compliance reviews on this issue. Targeting decisions, made preliminarily by the enforcement offices, are based on a wide range of criteria: the degree of disparities in the provision of educational services, complaints received, information about the progress (or lack of progress) which a recipient has attempted or made over time to address a particular civil rights issue, information received from recipients or interest groups, and reasoned judgment that an array of factors merits the use of resources in a proactive review of a recipient's compliance with civil rights obligations.

Secondly, this guidance is not intended to serve as a "cookie cutter" model for all testing decisions that are made in the field. There are many types of tests -- and this guidance is limited to "high stakes" tests, as defined in the guidance. It is not intended to apply to tests, for instance, that are used to measure student or institutional advancement but for which no high stakes consequences for individual students are tied. Moreover, as with other investigative guidance, it should be read as guidance for OCR lawyers and investigators -- nothing more, and nothing less. This means the Guidance should be read as an explanation of the legal and conceptual framework needed for understanding the issues raised by challenges to testing and assessment practices. As such, the model (and pragmatic) investigative questions should be viewed as a starting point for assisting our employees with questions of great complexity. They do not define the "floor" of what must be asked any more than they define the "ceiling" of what may be asked. Those decisions, fundamentally, are best left to the OCR investigators and lawyers who are dealing with the specific cases and who know the recipients best.

Finally, and importantly, this guidance is not intended to represent any intention on the part of OCR to hold itself out as "the expert" on testing issues. Far from it, and as the Guidance explicitly recognizes, the expert judgments regarding complex determinations, e.g., validity, must be left to the educational experts on that subject.

CONCLUSION

Guidance in this area should have immediate application for OCR investigations and have a positive impact on removing artificial barriers -- based upon race, national origin, or gender -- to educational opportunities and benefits.

INVESTIGATIVE GUIDANCE

I. Introduction

When tests are used to make educational decisions, they should be used to measure students' abilities, knowledge, or qualifications, regardless of race, national origin, or gender. Civil rights concerns arise when test uses do not satisfy this standard; the result is that equal opportunity may not be a reality for many students. This guidance outlines the requirements of Federal law prohibiting misuse of tests and other assessment procedures that result in discrimination based on race, national origin, or gender. It is designed to provide a general analytical framework under Title VI and Title IX for determining the proper use of tests in the educational context.

OCR staff are encouraged to use this guidance, along with the attached Compendium of Legal and Technical Resources, as a basis for framing investigations in which test scores are used as a basis for educational decisions. Materials located at Tab A outline the statistical framework for establishing disparate impact. Materials at Tab B provide a sample approach to the collection of information on the issue of educational necessity. Tab C provides a glossary of terms relating to test validity.

In evaluating a test or other assessment procedure, it is important to consider how the test is being used. In some cases, it may be used to make a certification or selection decision (e.g., admission to a school, awarding of a scholarship, or teacher certification). In other cases, it may be used to classify students (e.g., to identify students as needing special education or special language services or as gifted and talented). In both cases, the test may be designed and/or used as just one part of a multi-component assessment process. If so, its use should be evaluated in that context.

II. Jurisdiction

In all cases, OCR must first decide whether it has jurisdiction over claims involving discriminatory use of educational tests. Under the Civil Rights Restoration Act of 1987, OCR generally has institution-wide jurisdiction over a recipient of Federal funds.

If an institution receives Federal funds, requirements of Title VI and Title IX apply to all of the academic, athletic, and extracurricular programs of the institution, whether conducted in facilities of the recipient or elsewhere. Title VI and Title IX cover the uses of property that the recipient owns and the activities that the recipient sponsors. Title VI and Title IX cover these operations, whether the individuals involved in a given activity are students, faculty, employees, or other participants and outsiders.

Page 2 - Investigative Guidance

~~III. Disparate Impact Analysis~~

OCR may apply a disparate impact analysis to allegations involving discriminatory test use by recipients. Under this analysis the use of any educational test which has a significant disparate impact on members of any race, national origin, or gender group is discriminatory, and a violation of Title VI and/or Title IX, respectively, unless it is educationally necessary.

In applying a disparate impact analysis, OCR staff should address the following questions:

- A. Does the recipient's use of an educational test result in the significantly disproportionate denial of an educational benefit or opportunity to members of a particular race, national origin, or gender?
- B. If so, does the recipient have evidence that the test is valid and reliable under professionally accepted standards for the purpose for which the recipient has chosen to use it?
- C. If so, do there exist alternative forms of assessment which would substantially serve the recipient's stated purpose and are valid and reliable for that purpose, but which have less of a disparate impact on members of the race, national origin, or gender group?

Each question is discussed in more detail below. Where, based on the evidence obtained in an investigation, OCR finds that the use of a test or assessment procedure caused or contributed to a disparate impact on members of a particular race, national origin, or gender (the first question), and the test or procedure does not meet the legal standard of educational necessity (the second and/or third question(s)), OCR should conclude that there was discrimination in violation of Title VI or Title IX under this disparate impact analysis.

A. Establishing Disparate Impact

Under a disparate impact analysis, a recipient's use of an educational test that causes or contributes to a disproportionate denial of an educational benefit or opportunity to members of a particular race, national origin, or gender is sufficient information to indicate a possible failure of compliance which should be investigated further.

Addressing whether there is disparate impact requires a three-step statistical analysis. (Tab A.) OCR staff have access to statistical references, including a computer disk containing the "z" test. OCR staff also have access to published, commercially available statistical packages, such as SPSS.

Page 3 - Investigative Guidance

B. Establishing Educational Necessity

Once OCR has determined that there exists a possible failure of compliance through a disparate impact theory of discrimination, OCR will determine whether the recipient can establish/prove that the use of the test or assessment procedure is educationally necessary.² The use of a test or procedure which has a disparate impact on members of any race, national origin, or gender is discriminatory, and may be a violation of Title VI or Title IX, unless it is educationally necessary, i.e., it is valid and reliable, and there is no alternative with less disparate impact that still meets the recipient's educational needs.

In evaluating the validity and reliability of a test or assessment procedure, OCR will rely upon generally accepted professional standards such as those described in the Standards for Educational and Psychological Testing prepared by a joint committee of the American Psychological Association, the American Educational Research Association, and the National Council on Measurement in Education; the Code of Fair Testing Practices in Education prepared by the Joint Committee on Testing Practices; and the EEOC Guidelines on Employee Selection Procedures. All decisions as to whether a test or procedure has met professionally accepted standards will be made by experts.

A proper determination of educational necessity requires OCR to request information from the recipient concerning the testing and assessment procedure under review and the availability of suitable alternatives that have less disparate effect. During an administrative enforcement proceeding, the recipient has the burden of showing that the assessment process is valid and reliable, while OCR has the burden of showing that there are suitable assessment alternatives that have less disparate impact but still meet the recipient's educational needs.

² There are certain types of test misuse where a recipient will not be able to meet the standard of educational necessity. For example, OCR may find a violation of Title VI or Title IX if a test or other assessment procedure has a disparate impact and is clearly not being used for the purpose(s) for which it was designed. Such misuse includes the situation where a test is being used as the sole or principal criterion for making educational decisions and it was clearly not designed to be used as such. In such cases, no further analysis is required. If a recipient can show that a test or assessment procedure with a disparate impact is being used correctly, OCR must proceed with the educational necessity inquiry -- requesting additional information from the recipient about the validity and reliability of the test and, if necessary, exploring possible alternatives.

Page 4 - Investigative Guidance

1. Validation Standards³

Validation, generally, is a process of evaluating the degree to which a test: measures what it claims to measure; is administered in the correct manner; and leads to legitimate inferences. More specifically, a test or assessment procedure will be valid when:

- * it measures the construct (characteristic, property, skill, ability, capacity, or behavior) it was intended to measure;
- * it is administered in a correct and appropriate manner, with regard to: testing setting, testing procedure (including the qualifications of the test-giver and the manner in which the test is given), and tested sample of people (e.g., using a test validated for adults to assess children would be improper)⁴; and
- * the inferences drawn from the resulting data are correct and appropriate.⁵

A test may be valid for one educational purpose or population of students, but not valid for another. Thus, a decision by OCR regarding whether a test is valid under the circumstances of a particular case is an inherently fact- and case-specific decision.

³ The inclusion in the Guidance of a discussion of validation standards is intended to guide investigators in gathering information about available validation studies. Again, the discussion is not intended to establish new standards for determining validity, but to reflect existing professional standards. However, it should be noted that the Standards for Educational and Psychological Testing, which was published in 1985, is currently being revised. Also, the discussion does not address whether validity evidence is required for each race, national origin, and gender group that constitutes a significant portion or number of the recipient's test-taking population. This issue, when it arises, will be addressed on a case-by-case basis.

⁴ Improper test administration can also be an issue in disparate treatment, as discussed below.

⁵ Indeed, it may not be technically correct to refer to a test or assessment procedure as being valid. Rather, it is the inferences and interpretations drawn from the responses to the test or procedure that must be valid. However, for simplicity's sake, this guidance will use the more common approach of referring to the test or procedure as being valid for the purpose for which it is being used.

Page 5 - Investigative Guidance

There are various types of test validity evidence, including (among others) content validity, criterion validity (including concurrent validity and predictive validity), and construct validity.⁶

Criterion-related validity, the predictive form, is appropriate when a recipient is using test scores to predict students' future performance on a particular criterion or performance measure. Content validity and the concurrent validity form of criterion validity are relevant and appropriate when the educational concern is the assessment of current performance on a particular criterion or performance measure. Construct validity is an assessment of the extent to which the test is a measure of the particular construct or psychological concept (a characteristic, property, skill, ability, capacity or behavior) that the test is supposed to measure, and also, of the extent that the empirical results regarding the test are in conformity with existing theory concerning the construct.

To establish the criterion-related validity (predictive or concurrent) of a test the recipient should establish, through the use of empirical evidence, which is consistent with accepted professional standards, that the test scores are related to performance on the relevant criteria to a statistically significant degree, e.g., significant correlations. For example, a test which was designed to predict applicants' ability to earn a degree in mechanical engineering and used by a recipient to measure applicants' chances for success in a college mechanical engineering program, would be subject to predictive validity standards.

Content-related validity is implicated when a recipient is using a test to measure the acquisition of specific knowledge or academic skills. For example, a statewide proficiency test designed to demonstrate acquisition of educational information and/or skills and used as a condition for a diploma would be subject to an assessment of its content validity. To establish content validity a recipient should produce empirical evidence, consistent with accepted professional standards, of the degree to which the samples of items, tasks, or questions on a test are representative of the knowledge and skills being measured.

After determination of a test's content validity, it may be appropriate to inquire into the instructional validity related to the test. Instructional validity determines the degree to which a school provides instruction in the knowledge and skills measured by the test. A statewide proficiency test could also be subject to an instructional validity analysis.

⁶ Other types of validity are described in the Glossary at Tab C.

Page 6 - Investigative Guidance

Construct validity, which is the most general type of validity, is not determined in a single way or by one investigation. Rather it involves a series of studies and other procedures designed to determine whether an assessment instrument that purportedly measures a certain construct or variable is actually doing its job.

2. Establishing Validity

In appropriate cases, OCR staff will gather information relating to validation issues. Tab B provides guidance on the types of questions to ask and information to obtain regarding validity and the other aspects of educational necessity. Given the fact-specific nature of OCR's case work, these sample questions should be considered as starting points for appropriate inquiry. In most cases, these questions should be refined, modified, and supplemented based on the advice of testing and/or other education experts. Prior to requesting any validity evidence from a recipient, OCR staff should first find out if OCR already has any relevant documentation on the test or assessment procedure at issue. All decisions as to whether a test or procedure has met professionally accepted standards will be made by experts.

In evaluating the adequacy of the empirical evidence proffered by a recipient to establish that the use of a test or assessment procedure is educationally necessary, OCR will use the professionally accepted standards cited in this guidance and will additionally rely upon the recommendations of experts from within and outside the Department. The following guidelines should be considered when evaluating evidence of validity:

- a. No assumption of validity. The general reputation of a test, its author, or its publisher, or casual reports of its validity are not evidence of a test's validity. OCR will not assume that a test is valid based on a test's name or descriptive labels; promotional literature about the test; data regarding the frequency of a test's use; or testimonial statements and credentials of test publishers, consultants, or recipients which have previously used the test.
- b. A recipient may rely upon portions of a publisher's test manual as evidence of validity; however, a test manual is not presumptive evidence of validity. During the course of an investigation, there should be evidence of specific studies cited in the manual which show that the test is valid according to professionally accepted standards. These studies may be obtained

Page 7 - Investigative Guidance

By the recipient as a response to an OCR request for evidence of validation and need not be routinely maintained by the recipient.

b. Acceptable types of validity evidence.

Recipients may support the use of a test through validity studies of the same test conducted by the recipient, other schools, test publishers or distributors, or professional researchers. Such validity studies must show that the use of the test by the recipient is the professionally accepted equivalent to the use for which the test was validated.

c. Statistical relationships. As one part of the process of showing that a test or procedure is valid, the degree of relationship between test scores and performance criteria should be evaluated and determined. This may be done by recipients (or other researchers) using professionally accepted research and statistical procedures.

3. Establishing Reliability

Along with evidence of a test's validity, there should be evidence of a test's reliability over time. Evidence of reliability must conform to accepted professional standards. For example, where there are theoretical or empirical reasons for expecting reliabilities or standard errors of measurement to differ substantially for different populations, it is professionally accepted that the studies will include estimates for each major population for which the test is recommended. Moreover, reliability may be affected by the type of assessment procedure at issue, e.g., a standardized test versus a performance-based assessment.⁷

⁷ Performance-based assessment requires students to generate rather than choose a response. Students are required to actively accomplish complex and significant tasks, while bringing to bear prior knowledge, recent learning, and relevant skills to solve problems. Demonstrations, written or oral responses, journals and portfolios are examples of performance-based assessment. Herman, J.L., Schbacher, P.R., & Winters, L. (1992). A Practical Guide to Alternative Assessment. Alexandria, VA: Association for Supervision and Curriculum Development.

Page 8 - Investigative Guidance

4. ~~Alternative Testing and Assessment Procedures~~

Even if a recipient can show that a test or assessment procedure is valid and reliable, OCR may consider the recipient's continued use of the test or procedure in violation of OCR regulations if one or more testing instruments or procedures are available as an alternative to the challenged test or procedure, and if any alternative 1) substantially serves the educational purposes for which the test or procedure is used, 2) is valid and reliable for those purposes, and 3) would have a lesser disparate impact.

In evaluating the suitability of any alternatives, where the relevant method of validation is criterion-related, OCR should compare the predictive validity of those alternatives with the predictive validity of the challenged test or procedure. Where the use of a test or procedure offers only minor improvements in predictive validity over available alternatives, OCR should closely examine the test or procedure for evidence that it is necessary to achieve the recipient's purpose.⁸

C. Use of Educationally Necessary Tests or Procedures

1. Cutoff Scores.

In determining whether a test or procedure with a disparate impact is educationally necessary, OCR will look to how the test or procedure is actually used by the recipient. In some cases, a test or assessment procedure may be used without a specific passing or cutoff score. In other cases, a score may be set, either by the test developer or the test user. Standard 6.9 of the 1985 Standards for Educational and Psychological Tests states that "[w]hen a specific cut score is used to select, classify or certify test takers, the method and rationale for setting that cut score, including any technical analyses, should be presented in a manual or report." OCR will consider this information in determining whether a test or procedure is valid and reliable when used with that score.

In some cases, a recipient may use a higher score than that specified or recommended by the test developer. In such a case, it is critical to obtain the recipient's rationale for changing the score, particularly where a recipient's use of a higher score results in a greater disparate impact than would the lower score. It is not self-evident that a score of 75 will relate to better

⁸ It is good educational practice for recipients using tests or procedures that have disparate impact to inquire into alternative testing instruments or assessment procedures that would provide students with equal access to the recipient's programs and benefits.

Page 9 - Investigative Guidance

criterion performance⁹ than a score of 65 or 70. Because tests are often used to estimate criterion performance based on test performance, the recipient will have to show that incremental improvement in performance on the test above the cutoff score recommended by the test developer actually does translate into or predict better performance in school or on the job (e.g., if the test is used to predict success in a college mechanical engineering program, that the incremental improvement in test performance will actually translate into better performance in that program).

2. Affirmative Action

When a recipient has previously been found to have discriminated against persons on the basis of race, national origin, or gender or when the recipient is attempting to overcome the effects of conditions which resulted in limiting participation on those bases (e.g., attempting to enroll a more diverse student body), permissible affirmative action may include using test or assessment results in a different manner for members of a particular race, national origin, or gender.¹⁰

D. Remedies Under A Disparate Impact Analysis

Should a violation be identified, OCR will utilize a range of remedies to assist recipients in complying with Title VI and Title IX -- remedies which ensure equal access and promote educational excellence.

In any agreement with a recipient to remedy the effects of discriminatory test use, OCR should identify:

- 1) specific acts or steps the recipient will take to bring the recipient into compliance with the law;
- 2) the timetable for implementing each act or step; and
- 3) a specific timetable for submission of documentation.

⁹ Criterion performance is a measure of performance on some criterion, such as successful performance on the job or in an academic program.

¹⁰ Also see discussion of disparate treatment below, as to how affirmative action can be a legitimate nondiscriminatory reason for using a test or procedure differently based on race, gender, or national origin.

Page 10 - Investigative Guidance

All agreements should be crafted with a view toward effective monitoring, and permit flexibility to assist recipients in both complying with civil rights statutes and meeting their educational goals.

Examples of remedial efforts may include, but are not limited to, the following provisions. In an agreement for corrective action, OCR may wish to require further evidence of a test's or procedure's validity for a particular race, national origin, or gender, if the use of the test or procedure causes a disparate impact and is not, at the time of the investigation, fully supported by the required evidence of validity. Where appropriate, a corrective action agreement may include the continued use of the test or procedure in conjunction with other criteria. OCR may require a recipient that has not previously considered using additional testing instruments or assessment procedures as an alternative to the test or procedure that has been found to violate Title VI or Title IX to implement alternatives that 1) substantially serve the educational purposes for which the challenged test or procedure is used, 2) are valid and reliable for those purposes, and 3) have a lesser disparate impact. Additionally, there may be cases where OCR may also require complete suspension of the use of a test if a less discriminatory alternative is available, or if a recipient has shown bad faith in the use, or justification of the use, of a test. Also, where appropriate, a corrective action agreement may include providing students with learning opportunities to permit them to master material covered by a test.

IV. DIFFERENT TREATMENT ANALYSIS

If warranted by the nature and scope of the allegations or evidence, OCR will undertake a different treatment analysis, as described below, to determine whether the recipient administered a test or assessment procedure differently or used scores differently for students of a particular race, national origin, or gender, without a legitimate, nondiscriminatory reason.

Tests and assessment procedures must be administered and scores used in the same manner regardless of race, national origin, or gender. Even if a test or procedure is supported by sufficient evidence of educational necessity, a recipient may still be in violation of Title VI and/or Title IX if the test or procedure is administered differently or the scores are used differently for students of a particular race, national origin, or gender without a legitimate, nondiscriminatory reason.

A different treatment analysis can be pursued on a class-wide or individual basis. A class-wide pattern and practice approach will be more effective than an individual approach in identifying systemic problems. The pattern and practice case begins with a statistical demonstration similar to the one that begins a

Page 11 - Investigative Guidance

disparate impact case. However, if this disproportion is not the result of the application of a neutral policy, procedure, or practice, it may be evidence of a pattern and practice of different treatment and the analysis discussed below should be followed.

In applying the different treatment analysis, OCR staff should address the following questions:

1. In the administration of a test or procedure or in the use of scores for the granting or denial of an educational benefit or opportunity, did a recipient treat an individual or group differently than another individual or group, where they are similarly situated except for their race, national origin, or gender?
2. Did the circumstances of the test/procedure use provide a legitimate, nondiscriminatory reason for the different treatment?
3. Was the reason given by the recipient actually a pretext for discrimination?

Where, based on the evidence obtained in the investigation, OCR finds that a recipient administered a test or assessment procedure or used scores differently based on the race, national origin, or gender of test-takers, without a nondiscriminatory, nonpretextual reason, OCR may conclude there was discrimination in violation of Title VI or Title IX under the different treatment analysis.

On the other hand, if OCR finds that the reason for the different treatment was, e.g.,: 1) the provision of testing accommodations or auxiliary aids to qualified individuals with disabilities as required by Section 504 of the Rehabilitation Act of 1973 or Title II of the Americans with Disabilities Act of 1991; or 2) voluntary or remedial affirmative action undertaken in accordance with this guidance and the regulations that implement Title VI and Title IX, OCR may find that the recipient had a legitimate nondiscriminatory reason for the different treatment and may find no violation under this theory.

V. ESTABLISHING A VIOLATION OF TITLE VI BASED UPON A RECIPIENT'S USE OF A TEST OR ASSESSMENT PROCEDURE AFTER A FINDING THAT A RECIPIENT OPERATED A DUAL SYSTEM

The use of an educational test or assessment procedure may be a violation of Title VI if its use is a vestige of the previously segregated system -- i.e., it is a policy and practice traceable to the prior system. Additionally, the use of any test or procedure that, in reality, perpetuates the effects of previous discrimination may violate Title VI. Recipients that have operated dual systems and have not been declared unitary have an

Page 12 - Investigative Guidance

obligation to dismantle their prior de jure segregated systems. Such recipients have an obligation to identify, consider and implement less discriminatory criteria consistent with sound educational policy, to the extent practicable. Once it is established that a test or procedure is traceable to prior de jure segregation, the recipient must demonstrate either (1) that the test or procedure has no current segregative effects, or (2) that there are no less segregative alternatives to the test or procedure that are practicable and educationally sound.