Articles

Disparate Impact of Teacher Competency Testing on Minorities: Don’t Blame the Test-Takers — or the Tests

Michael A. Rebell*

Spurred by the recent spate of Commission reports that have decried the state of education in America,1 many states have recently enacted legislation to raise educational standards. By 1984, over half of the states had implemented some form of standardized testing requirement to assess the competence of prospective teachers. By 1988, nine more will join their ranks.2 In addition, a number of states have enacted legislation requiring incumbent teachers to prove their competence through standardized examinations.3

The trend toward standardized testing for teacher certification appears to be motivated by a widely held perception that many teacher candidates are ill-prepared to undertake the critical function of educ-

* B.A., Harvard College, 1965; L.L.B., Yale Law School, 1970; Partner, Rebell & Katzive, Attorneys, New York, NY; Visiting Lecturer in Law, Yale Law School. The author has represented plaintiffs in a number of cases challenging the validity of certification tests, and has also served as a legal consultant to several state education departments and to a test development company. The author would like to thank Dr. Bernard R. Gifford, Dean of the Graduate School of Education, University of California at Berkeley, and former Deputy Chancellor, City School District of the City of New York, for discussing many of the themes set forth in this article with him over the past year. Appreciation is also extended to William Gorth, Merle McClung, William Mehrens, Herbert Teitelbaum, and Carol Ziegler for their comments and suggestions.


2. M. Goertz, R. Ekstrom & R. Coley, The Impact of State Policy on Entrance Into the Teaching Profession at 9 (Final Report, National Institute of Education Grant No. GA3-0073, 1984). The tests vary from state to state according to when they must be taken, i.e., prior to, during, or after entry into a teacher training program. Also, some assess only basic skills or knowledge; others assess more extensive subject matter mastery or classroom skills.

cating the nation's youth. A good deal of evidence substantiates this perception. The lowered relative salary scales and diminished prestige of the teaching profession have reduced both the number and the qualifications of applicants to teacher training programs. These factors have undermined the traditional system in which graduates of state-approved teacher training programs were presumed competent and were automatically granted state certification. In the wake of a perceived crisis, policymakers have begun to implement a number of educational reforms.

The long-range solution to the problem of teacher competence clearly requires increasing teacher salaries, enhancing the status of the profession and raising the caliber of teacher training programs' academic offerings. Some of these reforms, most notably salary increases, have already begun to be instituted. While long-range measures are being considered, however, the strong political demand for immediate assurance of competence in the classroom is being placated by implementing standardized testing requirements. These tests are politically appealing because they provide a quick, relatively inexpensive mechanism for weeding out unqualified teacher candidates and because they promise to impose objective external standards.

Standardized testing of teachers, however, raises a number of significant questions: Can a battery of pen and paper tests truly assess the full range of skills needed to perform effectively in the classroom? If a test purports to cover only knowledge of minimum subject matter content, is it a fair criterion for certification, or should college grade point averages or performance evaluations also enter into the equation? If tests are generally valid, are they differentially biased against particular minority groups? Most of the recent political and legal controversy has focused on the last of these questions because of the highly adverse impact the tests have had on minorities. That thousands of black and Hispanic students who have prepared for teaching careers are being denied entrance certification to a profession in which minorities are already underrepresented is a

---

4. The number of college women majoring in education declined from 19 percent in 1972 to 10 percent in 1984; the number of men majoring in education declined from 6 percent to 3 percent during the same time period. P. Garcia, A STUDY ON TEACHER COMPETENCY TESTING AND TEST VALIDITY WITH IMPLICATIONS FOR MINORITIES 7 (National Institute of Education Grant No. G-85-0004, 1985). Teachers-in-training ranked fourteenth out of sixteen occupational groups on SAT verbal scores and fifteenth out of sixteen on quantitative scores. Kirst, RENEWING THE TEACHING PROFESSION, STANFORD MAG., Spring 1985, at 52. Several prestigious institutions, including Yale, Harvard, Reed and Duke, dropped their undergraduate teacher training programs in the 1970's. Id. at 53.
Teacher Competency Testing

serious and pressing social issue. Needless to say, the disparate impact of teacher certification tests on minority candidates has generated a substantial amount of litigation.

In this article, I will describe the extent of the disparate impact suffered by minority candidates and explore the general legal standards for validating tests which have an adverse impact on minority applicants. Under current case law, most certification tests would pass muster despite their impact on minorities. I will here trace the development of two new legal theories which, however, if they are widely accepted by the courts, might invalidate many of the tests.

I will argue that these new legal theories exacerbate rather than solve the problem. Low minority pass rates do not reflect either bias in the tests or lack of aptitude in the candidates, but are a result of a substantial underlying deficiency in the candidates' academic preparation. Especially in the Deep South, where most of the disparate impact has been felt, educational shortfalls stem not only from teacher training programs but also from the entire legacy of inferior, segregated schooling. To certify a wave of well-intentioned but ill-prepared teacher candidates would only perpetuate the cycle of unequal educational opportunity.

Accordingly, the objective standards reflected in the tests must be maintained as critical accountability measures in order to compel long-range reform of educational offerings to university students. In the short run, motivated candidates who fail on a first attempt should be provided with intensive remedial opportunities to assist them in improving their skills so that they eventually obtain certification. These recommendations will be considered at greater length below.

I. Disparate Impact and Test Validation

The extent of the disparate impact of teacher certification examinations on minority candidates is evident from the results of the first administration of the California Basic Educational Skills Test in 1983. While 76 percent of the white candidates passed, the pass rate for minority candidates was markedly lower. Only 26 percent of blacks, 38 percent of Mexican Americans and 50 percent of Asian Americans passed the test.\textsuperscript{5} Recent teacher certification tests in the Deep South produced similar outcomes. Results of the 1983 National Teacher Examination (NTE) in Louisiana and the 1983 Flor-

\textsuperscript{5} P. Garcia, supra note 4, at Appendix C.
ida Teacher Certification Exam show respectively a 78 percent and 90 percent pass rates for whites but only a 15 percent and 35 percent pass rates for blacks. If these trends continue, "minority representation in the national teaching force could be reduced to less than 5 percent by 1990."7

Part of the problem revealed by these disturbing statistics can be traced to heavy demands on the available talent pool. As noted above, minority students' interest in teaching has declined significantly in the past fifteen years as new career opportunities outside education have become available to them for the first time. Nevertheless, the scope of the disparate pass rates, and their impact on the lives of thousands of individuals whose entry into their chosen career is blocked off, cannot be ignored. The results of recent litigation guarantee that the problem will not go unnoticed. Courts in Alabama, Arkansas, and Texas are currently considering these issues, and other suits are sure to follow.

Challenges to teacher certification tests are not new. In the 1970s, a wave of litigation in the Deep South challenged the use of the NTE as a certification device. The NTE is an examination which was created by the Educational Testing Service (ETS) to assess students' knowledge of the typical curricula taught at teacher training institutions throughout the country. ETS did not recommend or authorize the use of the NTE as a certification device in any particular state since the exam had not been validated for this purpose. Under these conditions, courts did not hesitate to enjoin use of the NTE as a certification or job retention requirement where it was shown to have the effect of denying continued employment to black teachers in newly desegregated school systems.9 The current generation of testing litigation, however, raises more difficult legal and policy issues. Most of the recently implemented testing requirements have been adopted as part of a broad education reform program, rather

---

6. Id.
Teacher Competency Testing

than for clearly discriminatory reasons. Moreover, most tests have been professionally developed and extensive efforts have been undertaken to validate them for the specific purposes for which they are being used.10

Of course, despite stated benign purposes, the fact that a state education department adopts a testing program which is likely to have a substantial disparate impact on minority candidates can raise questions concerning the motivation behind the reforms, especially if the state in question is one with a history of de jure segregation. Under such circumstances, it is possible that a state might have adopted all or part of a professionally developed testing program as a sophisticated device with which to perpetuate or even extend past discriminatory hiring practices.

On the other hand, if the remaining vestiges of racial segregation are to be overcome, it is necessary to break the cycle of failure that has caused low minority achievement rates on standardized tests. Allowing ill-prepared individuals to teach youngsters in the predominantly minority schools to which the less qualified teachers tend to be assigned will perpetuate historical patterns of unequal opportunities indefinitely.11 In order to ensure that the next generation of minority students receives competent instruction, high standards for teacher competency must be maintained.12 This is so despite the fact that insistence upon high standards has an unfortunate immediate impact on present minority candidates who, often through no fault of their own, were not properly trained or given fair opportunities to prepare for a demanding professional career.

10. Two types of teacher competency testing examinations are currently used. First is the NTE, which although it remains a standardized national examination of knowledge and skills taught in teacher training institutions is now generally validated by particular state or local school districts to insure that it reflects knowledge and skills needed in that particular setting. The second consists of "customized" examinations prepared by National Evaluation Systems, Inc. (NES) of Amherst, Massachusetts. The NES tests are based on job analyses conducted in each state to reflect local teaching requirements directly. See Flippo, Teacher Certification Testing Across the United States and A Consideration of Some of the Issues 6, Tables 1 and 2 (Paper presented at annual meeting of American Educational Research Association, Chicago, 1985).

11. "I assume that the reason minority applicants fare worse on the test than Whites is that they themselves are victims of inferior schooling." Washington Post columnist William Raspberry, quoted in Gifford, supra note 8, at 57. Gifford notes that major differentials in test scores appear to be related to class bias factors as well as to race bias factors. He indicates that the College Board’s Profiles, College-Bound Seniors, 1983, shows that "[T]he relationship between family income and test scores is highly significant. While not as high, the relationship between level of parental education and SAT scores of high school seniors is also very substantial." Id. at 56.

Standardized tests are criticized because they cannot measure personal warmth and caring, the ability to maintain order, and other qualities important for effective teaching. Well-constructed tests, however, can measure a candidate's basic knowledge of the subject matter at hand, knowledge which is indisputably a *sine qua non* for competent teaching. It may eventually become possible to develop more effective tools for objectively measuring the broad range of skills, personal traits, instructional techniques and subject matter knowledge that is required. That such broad-based evaluative techniques are not currently available does not mean that the more limited objective measures of minimum subject matter competence which we do have should not be used. As Nathan Glazer has said: "It is easy to attack tests: the question is whether there is an alternative. Tests correlate roughly with some kind of ability... To attempt to introduce other qualities means to depend on the uncertain outcomes of interviews and other kinds of qualitative assessment."\(^{13}\)

Teacher competency testing is not the only context in which standardized tests have raised these issues. Such tests are used for many purposes and have resulted in disparate impacts on minorities in many other situations. The courts have therefore had to wrestle extensively in recent years with the issue of what level of test validation should be required when minorities are detrimentally affected. For the most part, the courts have scrutinized the tests under Title VII of the 1964 Civil Rights Act, the major federal statute prohibiting discrimination in employment.\(^{14}\) The statute allows an employer to use "a professionally developed ability test" only if it is not "designed, intended or used" for discriminatory purposes.\(^{15}\)

The requirements for test validation developed by the courts and the regulatory agencies in response to Title VII tend to be accepted as the general operative standards for all employment-related tests. It is not clear, however, whether all licensing and certification examinations which are prerequisites for employment are "employment

---

\(^{13}\) Id. at 227. Standardized minimum competency tests also provide a more objective assessment method than the unvalidated, often politically sensitive, traditional "program approval" certification approach. See Mehrens, *Validity Issues in Teacher Competency Tests* 8 (Prepared for the Institute for Student Assessment and Evaluation, University of Florida, 1986). Because teacher competency tests can provide some reliable information about the basic competence of applicants but cannot provide a full rating of all pedagogical skills, test results should be used as a threshold eligibility requirement, and not as a method to rank applicants.


tests” under Title VII. Whether or not Title VII technically applies to teacher certification tests, well-developed minimum subject competency tests have little difficulty meeting its validation standards.

The Title VII validation standards are set forth in guidelines that have been issued by the Federal Equal Employment Opportunity Commission (EEOC). These guidelines were cited approvingly and applied by the Supreme Court in *Griggs v. Duke Power Company* and a series of later decisions. Validation under the EEOC Guidelines falls into several categories, the most significant of which are “criterion-related validation” and “content validation.” Criterion-related validation generally requires an employer using a test to demonstrate that those examinees who pass the test or receive higher grades on it perform better on the job than those who fail or get lower grades. In other words, it usually requires empirical evidence that the test actually predicts competence on the job. In the teacher certification context, criterion-related validation would require a showing that those exhibiting greater knowledge and skills

16. Compare, e.g., Tyler v. Vickery 517 F.2d 1089 (5th Cir. 1975) and Woodard v. Virginia Bd. of Bar Examiners, 598 F.2d 1345, 1346 (4th Cir. 1979) (State Board of Bar Examiners is not an “employer” within the meaning of Title VII) and NOW v. Waterfront Comm’n of New York Harbor, 468 F. Supp. 317 (S.D.N.Y. 1979) (State waterfront licensing agency is not an employer) with Vanguard Justice Soc’y v. Hughes, 471 F. Supp. 670, 696 (D. Md. 1979) (City civil service commission is the “employer” of police) and Puntolillo v. New Hampshire Racing Commission, 375 F. Supp. 1089 (D. N.H. 1974) (State racing commission is the “employer” of driver-trainers) and United States v. North Carolina, No. 4476 (E.D.N.C. June 23, 1982) (State Board of Education is the “employer” of local school district teachers).

17. The Guidelines, which have been endorsed by the Justice Department, the Department of Labor and other federal agencies, are officially known as the Uniform Guidelines on Employee Selection Procedures (1978). They are codified at 29 C.F.R. Part 1607. The Guidelines specifically refer to the test validation standards contained in *Am. Psychological Ass’n, Standards for Educational and Psychological Testing*, (rev. ed. 1985).

18. 401 U.S. 424 (1971). The main holding in *Griggs* was that Title VII proscribes not only intentional discrimination, but also the use of tests which are neutral in intent but have a disparate impact on minorities. The Court held that tests resulting in a disparate impact may be used only if they are shown to be properly validated.


20. The Guidelines also refer to a third category, “construct validation,” which requires a showing that an examination measures identifiable traits, characteristics or “constructs” that have been shown to be important to successful performance on the job. Construct validity is, however, less significant in actual practice and its application is often difficult to distinguish from content validation. *See, e.g.*, Guardians Ass’n of New York City v. Civil Service Comm’n, 630 F.2d 79 (2d Cir. 1980).
on the tests actually perform better in the classroom. It is generally acknowledged that the current state of the art of test measurement ("psychometrics") cannot achieve accurate predictive correlations of this type.

The alternative validation approach, "content validation," requires the employer to show that the content of a test is reasonably related to the knowledge and skills required for effective performance on the job. In the teacher testing context, minimal subject matter competency tests which include some, if not all, of the knowledge needed for effective teaching would be assessed under a content validation standard. Although earlier versions of the EEOC Guidelines contained an explicit preference for criterion-related validation, under the current Guidelines either validation procedure is considered equally acceptable. Therefore, content-based tests, if properly developed, would be acceptable under the Guidelines even if they had a substantial adverse impact on minority applicants.

In the early years of enforcement of Title VII and the EEOC Guidelines, many employment tests which had a disparate impact on minorities were invalidated by the courts. Those cases are, however, largely irrelevant to the current crop of teacher certification tests. Most of the invalidated tests were ad hoc selection devices. No serious attempts had been made to analyze the requirements of the job and correlate test content with them.21 As companies and public employers became more aware of the legal requirements and more sophisticated in their use of testing devices, the trend shifted dramatically; after 1976, tests were validated in many more cases.22

The U.S. Supreme Court's 1976 decision in Washington v. Davis 23 further influenced the trend toward upholding the use of employment tests, even those with a disparate impact. There, the Court upheld the use of a verbal ability test for police recruits despite its disparate impact on minority applicants and its failure to meet a technical requirement of the EEOC Guidelines. Although the Court found that the EEOC Guidelines were not directly applicable in Davis, it seemed willing to accept a lower standard than that of the EEOC because the case involved a professionally developed testing program that had been adopted in good faith by a public em-

---

21. See, Rebell & Block, supra note 19, at 13. (Plaintiffs prevailed in fifty-six of seventy Title VII cases decided prior to 1976.)
22. Id. at 22. (Plaintiffs prevailed in eighteen of thirty-seven cases decided after 1976.)
Teacher Competency Testing

ployer.24 This attitude was further evidenced by the Supreme Court’s later summary affirmance, in United States v. South Carolina, of a three-judge district court’s approval of the South Carolina NTE despite a strong disparate impact on minority candidates.25

In sum, the EEOC Guidelines’ inclusion of content validation techniques as an alternative methodology, together with the lower standard applied by the Supreme Court where public employers have made good faith attempts to use professionally developed tests, suggests that properly developed teacher certification tests are likely to be upheld by the courts even if they have substantial disparate impact on minority applicants. Moreover, since these are “licensing” tests and are imbued with a significant public policy purpose, protecting the public from incompetent practitioners, the courts are likely to be even more deferential to defendants than they are in the strict employment context.26

24. In Washington v. Davis, Title VII technically did not apply because the case was filed prior to 1972, when Title VII was extended to cover state and municipal employees. Strictly applied, although a positive correlation had been shown between success on the test and performance in the police academy training program, 426 U.S. at 246, the Guidelines would have required a further showing of a correlation between the requirements of the training program and success on the job. See 29 C.F.R. § 1607.4(c) (1975). The Court’s relaxed attitude was partly due to the fact that the Washington, D.C. police force had made significant strides in recruiting minority applicants. Significantly, even though the Court technically was not interpreting Title VII or the EEOC Guidelines, the majority decision went out of its way to indicate that the job-relatedness requirement of Title VII, if applicable, would not have been interpreted to require a different result: “Nor is the conclusion foreclosed by either Griggs or Albemarle Company v. Moody, 422 U.S. 405 (1975); and it seems to us the much more sensible construction of the job-relatedness requirement.” Id. at 250-51.

25. United States v. South Carolina, 445 F. Supp. 1094 (D.S.C. 1977), aff’d, 434 U.S. 1026 (1978). In order to validate the NTE for local certification purposes, ETS had assembled local educators in South Carolina who opined that the content of the NTE, which was created to test knowledge of subject matter taught in teacher training institutions nationwide, reflected the particular curriculum being taught in the teacher training institutions in that state. 445 F. Supp. at 1112. Plaintiffs argued that the job-relatedness requirements of the EEOC Guidelines required not a correlation with training course content, but a correlation with the skills and knowledge needed for effective performance as a classroom teacher on the job. See id. at 1108 n. 13. The three-judge panel, relying on Davis, upheld the training course validation. Justice White, the author of the Davis opinion, strongly dissented from the Supreme Court’s summary affirmance:

... Washington v. Davis, in this respect, held only that the test ... which sought to ascertain whether the applicant had the minimum communication skills necessary to understand the offerings in a police training course, could be used to measure eligibility to enter that program. The case did not hold that a training course, the completion of which is required for employment, need not itself be validated in terms of job-relatedness. Nor did it hold that a test that a job applicant must pass and that is designed to indicate his mastery of the materials or skills taught in the training course, can be validated without reference to the job.

434 U.S. at 1027.

Plaintiffs in the current teacher competency testing cases, unhappy with the existing standards, have developed two significant new legal theories which would hold state certification boards to higher validation requirements than the Supreme Court has endorsed. The first, "instructional validation," demands that test content be geared to the curricula of teacher training colleges. The second would require the deletion from examinations of questions on which minorities do not score well. If these theories are accepted by the courts, either teacher competency tests will have to be abandoned or fundamental alterations will have to be made in test standards to accommodate the present level of minority candidates' preparation. Neither of these approaches is desirable as a matter of policy or justified as a matter of law.

II. Instructional Validation

A. Student Competency Testing

The Supreme Court has not explored test validation concepts since its major rulings a decade ago. Thus, the lower federal courts have been called upon to apply — and possibly extend — the Title VII test validation standards without specific guidance from the Court. One area in which the lower federal courts have extended Title VII concepts, even though technically the statute does not apply, is student competency testing. In Debra P. v. Turlington, a student testing case, the court of appeals developed the concept of "instructional validation." Because this concept may become a major issue in the teacher certification area, Debra P. must be examined in some detail.

The Debra P. litigation was part of a concerted effort by the Harvard Center on Law and Education and other advocacy groups to challenge minimum competency statutes which have been adopted in recent years by more than two-thirds of the states.

of the qualifications required of doctors must depend primarily upon the judgment of the State as to their necessity."); Douglas v. Noble, 261 U.S. 165 (1923) (States may delegate broad authority to administrative agencies to examine the qualifications of dentists); Smith v. California, 336 F.2d 530, 534 (9th Cir. 1964) ("The principles stated in the Dent case have been widely applied by all the States in a great variety of professions . . . "); Goldfarb v. Virginia State Bar, 421 U.S. 773, 792 (1975) ("[T]he States . . . have broad power to establish standards for licensing practitioners and regulating the practice of professions.")


28. 730 F.2d at 1407-09.

29. Thirty-seven states have enacted minimum competency legislation. Logar, Mini-
Teacher Competency Testing

Minimum competency testing involves the use of standardized testing instruments to assess student mastery of "basic skills," usually reading, writing and mathematics. While standardized achievement tests have been used in secondary education for years, the minimum competency testing programs are new in that they are being implemented on a statewide basis to prescribe remedial help or deny high school diplomas to students who fail the tests. Minimum competency tests (MCTs) tend to have a substantial disparate impact on minority students.30

At issue in Debra P. was the Florida "functional literacy examination," which had had a substantial adverse impact on minority students. The first time this test was administered, 78 percent of the black students but only 25 percent of the white students failed one or both sections.31 In light of these statistics, the district court closely examined whether the test met the validation requirements.32

Because of Florida's long history of de jure school segregation, the court paid special attention to the plaintiff's allegations of intentional discrimination. It rejected the contention that the State Commissioner's knowledge of the effects of the test on black school children necessarily constituted intentional discrimination. On the
contrary, the court applauded Florida’s motives in implementing a testing program to raise educational standards:

The legitimate interest in implementing a test to evaluate the established state-wide objectives is obvious. The minimal objectives established could be continually upgraded and the test could be utilized not only to gauge achievement, but also to identify deficiencies for the purpose of remediation.33

At the same time, however, the court also concluded that:

The timing of the program must be questioned to some extent because it sacrifices through the diploma sanction a large percentage of black twelfth grade students in the rush to implement the legislative mandate.34

Accordingly, the court enjoined Florida from requiring passage of the examination as a requirement for graduation for a period of four years.35

On appeal, the court accepted the lower court’s basic findings and affirmed its holding that the test items were not biased, but went beyond the EEOC test validation concepts upon which the district court had relied. It held that an additional requirement should be imposed upon student competency tests, namely a showing of “instructional validity.”36 Specifically, the state was required to prove that the “test covered things actually taught in the classrooms.”37 Because the lower court had not considered whether the “instructional validity” of the test had been established, the case was remanded.

The court of appeals’ application of the doctrine of instructional validity in this decision has generated substantial interest in psychometric circles.38 There is a widespread assumption that instruc-

33. 474 F. Supp. at 254.
34. Id.
35. Id. at 269.
36. The Court used the term “curricular validity,” but the concept it had in mind clearly was closer to “instructional validity.” McClung, an attorney for the plaintiffs in Debra P., apparently was the first to discuss these concepts extensively. He defined “instructional validity” as a measure of “whether or not the school district’s stated objectives were translated into topics actually taught in the district’s classrooms.” By way of contrast, the term “curriculum validity,” as defined by McClung, was a “measure of how well test items represent the objectives of the curriculum to which the test takers have been exposed.” See McClung, Competency Testing: Potential for Discrimination, Clearinghouse Rev., Sept. 1977, at 439, 446. In order to avoid confusion in terms, the validation requirement articulated by the Court in Debra P. will be referred to throughout this article as “instructional validity.”
37. 644 F.2d at 405.
38. Standard 8.7 of the newly adopted Standards for Educational and Psychological Testing, supra note 17, states:

When a test is used to make decisions about student promotion and graduation,
Teacher Competency Testing

tional validity has become a judicial requirement for student competency tests and perhaps for tests in related fields such as teacher certification as well. This assumption, however, overlooks the difficulty encountered by the district court when it attempted to implement the instructional validation standard. Practical and conceptual problems, clearly revealed in the record of the remand proceedings, led the district and appeals courts to become markedly more circumspect. The implications of the courts' post-remand consideration of the issue have not been sufficiently appreciated.

Although there is an obvious commonsense appeal to the notion that students should not be penalized for failing to test well on subject matter they have not been taught, the Debra P. remand hearing shows that any attempt to determine what individual students have actually been taught is impossible. In order to establish that every student in the state had had a fair opportunity to learn each of the many subjects covered by an examination, the practices of every school district, perhaps even of every school and classroom, in the state over the twelve year span of a public school education would have to be analyzed. Arguably, the need to demonstrate that a fair opportunity had been provided to each student could be sidestepped by aggregating experiences in order to show that the concepts covered on the tests were taught consistently throughout the state. This would pose serious problems in terms of individual rights, however. Moreover, limited review, even if doctrinally acceptable, would be an unmanageable enterprise that would intrude upon ongoing educational operations. The programs of hundreds of school districts would have to be assessed over extended periods of time. Educators are also not likely to welcome such an undertaking because "attempts to assess [instructional] validity on a statewide basis [could] lead to debilitating bureaucracy, costly adminis-

---


40. To be fully equitable, an additional showing of how well each individual child was taught, and how much or how recently, might also be required. See, e.g., Debra P. v. Turlington, 654 F.2d 1079, 1083 (on petition for rehearing en banc, Tjoflat, J., dissenting).

41. Methods proposed for accomplishing this include techniques for classroom observations, teacher self-reports, student self-reports, and development of individual pupil cumulative record cards. These issues are discussed generally in the articles collected in THE COURTS, VALIDITY AND COMPETENCY TESTING (G. Madaus ed. 1983). See also McClung, supra note 30, at 705-08.
The Debra P. remand decision brought these practical problems into sharp focus. In response to the court of appeals' decision, Florida retained a consulting firm to undertake an extensive four-part validation study. A teacher survey was sent to 65,000 teachers. A detailed survey of all of Florida's school districts and four university laboratories was taken. Site visit teams were sent to each of the school districts to follow up on the surveys. Thousands of student surveys were completed. Despite this comprehensive range of data, the plaintiffs argued that the validation study was fundamentally flawed because the survey was constructed to invite positive responses from the teachers. They also asserted that the survey was deficient because it covered only one rather than all twelve years of the students' education, and because it provided insufficient evidence of what actually happens in the classroom.

The district court's response revealed its frustration with instructional validation:

[A]bsent viewing a videotape of every student's school career, how can we know what really happened to each child? Even assuming that such videotapes were available, how could this Court decide, in constitutional terms, which students received appropriate instruction and which did not? Suppose that there is one student who never encountered a teacher who taught the SSAT-II skills, or a teacher who taught the skills well, should the entire test be declared invalid? What if the number of students were 3,000 rather than 1?43

In light of these difficulties, the district court abandoned the strict requirements of instructional validation and instead upheld the examination on a significantly more limited basis, a requirement of "curricular validity."44 The court explained, "[w]hat is required is that the skills be included in the official curriculum and that the majority of the teachers recognize them as being something they should teach."45 In other words, as long as the curriculum included the basic objectives and the teachers were aware of them, the court

44. As indicated supra note 36, although the Fifth Circuit's 1981 decision had used the term "curricular validity," it clearly meant to impose a requirement synonymous with "instructional validity" since it required a showing that test objectives "covered things actually taught in the classrooms." The court's confusion about the terms is in itself an indication of the conceptual complexities involved.
45. 564 F. Supp. at 186.
Teacher Competency Testing

would not insist on the more stringent instructional validity requirement that students had actually received instruction in each of the requisite objectives.

When the appeals court affirmed this holding, it also seemed less committed to instructional validity: "The experts conceded that there are no accepted educational standards for determining whether a test is instructionally valid." While the decision affirmed the district court's findings of fact as not "clearly erroneous," it was carefully qualified. It specifically eschewed the difficult task of establishing a clear legal standard. In sum, the court of appeals appeared to have studied the complex record of the remand proceedings below, realized the difficulties involved, and limited rather than extended the lower court holding.

B. Teacher Competency Testing

Despite the difficulties experienced by the courts in attempting to implement "instructional validation" in the student testing context, the concept has been widely invoked by plaintiffs challenging teacher certification tests. Instructional validation here would require that certification examinations correlate with the curriculum at the teacher training institutions, not to the knowledge and skills needed for competent teacher performance in the classroom. The
notion that candidates for teacher certification should be tested only on material to which they were exposed at their training institutions is appealing. Nevertheless, abstract equity must be weighed against the virtual impossibility of proving what was taught at each of the many teacher training institutions in every state. (Instructional validation could even require states to consider the curricula of out of state institutions candidates had attended.) Moreover, the rights of the public school students who will ultimately be taught by the candidates must be considered. The equity issue is problematic because the interests of teaching candidates must be weighed against the equal, arguably greater, interests of future minority students who are entitled to be taught by competent teachers.

The analogy between the student testing in Debra P. and the current teacher testing situation is even more tenuous when considered in light of applicable legal doctrines. Debra P. was based in large part on the due process notion of the "legitimate expectation" of high school students that they would receive a diploma if they passed all their courses and otherwise met the standards which were in effect prior to the implementation of the competency testing requirement. Candidates for teacher certification do not have such a "legitimate expectation" because simply attending school for four years does not entitle them to a teaching license. The concept of state credentialling is based squarely on the propositions that the public must be protected against incompetent practitioners and that no one has a right to enter the profession without demonstrating the requisite degree of competence.49

49. See United States v. Lulac, No. 85-2579, slip op. (5th Cir. July 2, 1986). Note, supra note 19, listed a number of other arguments for distinguishing the student testing situation in Debra P. from the teacher certification as follows:

First, enrollment in teacher training programs is not compulsory, unlike school attendance for children. Second, attendance in teacher training programs is relatively short term compared to the twelve years of attendance required for receipt of a high school diploma. Third, failure to obtain teacher certification does not preclude most other forms of gainful employment, unlike possession of a high school diploma, which today is a prerequisite to many forms of gainful employment. Id. at 502.

The Supreme Court upheld validation based on the local teacher training college curriculum in United States v. South Carolina, 434 U.S. 1026 (1978), supra note 25, but did so on a record which offered as the sole alternative to objective testing a discretionary approval procedure that the Court considered highly undesirable. See 445 F. Supp. at 1115-1116. If the psychometric profession can now provide a job-related certification examination as an acceptable alternative, courts should endorse this preferred approach. In this context, South Carolina would not provide precedent for adding an additional, and inconsistent, instructional validation requirement to an otherwise acceptable job-related examination. Cf. Ensley Branch of NAACP v. Seibels, 616 F.2d 812, 819 n.17

390
Teacher Competency Testing

In short, because of tenuous doctrinal support and serious practical proof problems, there is little basis for extending instructional validation to teacher certification. Moreover, it is doubtful that the courts will do so.\(^{50}\) Even without any direct mandate, however, instructional validity has begun to influence the development of teacher certification testing. State education departments are including an instructional validation component in their test construction procedures in order to avoid possible legal challenges. This undermines the purpose of teacher competency testing and content validation standards which base certification on demonstrated classroom competence rather than on familiarity with the often substandard teacher training curricula. The plight of teacher certification candidates who have attended training institutions that do not adequately prepare them for their chosen professions is real and should be addressed. Their needs would be better met, however, by providing them with greater opportunity to prepare for the examination.

III. Following the Golden Rule: The Item Bias Approach

There is another emerging approach to competency testing which threatens to undermine the maintenance of competency standards by teacher certification tests. Known as the "Golden Rule" approach, this new concept could have serious adverse effects on the integrity of testing.

Under Title VII and the EEOC Guidelines, employment selection tests which have an adverse impact on minorities may be adopted if the employer can show that reasonable psychometric methods were

---

(5th Cir. 1980), cert. denied, 449 U.S. 1061 (training program validation suitable only for tests of minimum reading and verbal skills).

50. Judge Justice recently referred to Debra P. in partially granting plaintiff's motion for a preliminary injunction to limit the use of the Texas Pre-Professional Skills Test (PPST) for entry into a teacher preparation program, holding that the state had to demonstrate "that students in Texas had been taught the materials covered in the PPST." United States v. Texas, 628 F. Supp 304, 320 (E.D. Tex. 1985). This decision was, however, reversed by the Fifth Circuit, in a decision which strongly emphasized the state's "interest in ensuring teacher competency". United States v. Lulac, No. 85-2579, slip op. at 7012 (5th Cir. July 2, 1986). The Fifth Circuit held that a preliminary injunction should not be issued before there had been a full trial on the merits of the complex test validation issues raised by the case. Speaking directly to the public policy concerns discussed in this article, the court also stated:

In administering its higher education systems, even a state that formerly practiced de jure segregation has no constitutional or statutory obligation to suspend or lower valid academic standards to accommodate high school students who may be ill-prepared because of prior constitutional violations by its local and elementary school systems.

Lulac, slip op. at 7015.
used in the tests' construction. In other words, in the absence of a showing of intentional discrimination, adherence to the validation requirements set forth in the Guidelines establishes a presumption that the adverse impact did not result from discriminatory practices.  

Even where sound validation practice has been followed, however, an employer has an obligation to consider the use of suitable alternative selection procedures if another method with less adverse impact on minorities exists and is "substantially equally valid for a given purpose." This provision is rarely applied because few alternative testing methods, especially for subject matter competence, have been developed by the testing industry. Virtually all standardized certification tests, whatever the form of their validation, have tended to have a substantial adverse impact on minority candidates.

Recently, however, certain new techniques for potentially reducing adverse impact on minorities have been proposed to the courts. These techniques eliminate specific examination questions ("items") on which minorities fare poorly. This approach does not take into account either the overall structure of the test or the basic constructs being tested. It focuses instead on the fact that certain test items in an otherwise acceptable test may prove more difficult for one group than for another. Elimination of "biased" items and the substitution of alternative items covering the same basic content is supposed to provide equity to all test-takers while assuring an objective measure of relative or absolute competence.

The item bias approach received significant support when it was incorporated into a consent decree in Golden Rule Insurance Company v. Washburn, a decree into which ETS entered. This protracted

51. "[N]othing in these guidelines is intended or should be interpreted as discouraging the use of a selection procedure for the purpose of determining qualifications or for the purpose of selection on the basis of relative qualifications, if the selection procedure has been validated in accord with these guidelines for each such purpose for which it is to be used." 29 C.F.R. § 1607.2(c) (1986).

52. 29 C.F.R. § 1607.3(b) (1986).

53. See, e.g., REPORT OF THE COMMITTEE ON ABILITY TESTING, NATIONAL RESEARCH COUNCIL, ABILITY TESTING: USES, CONSEQUENCES, AND CONTROVERSIES 18 (1982) ("certain social groups tend, as groups, to score consistently lower on the average than more advantaged groups . . ."); Clady v. County of Los Angeles, 770 F.2d 1421, 1432-33 (9th Cir. 1985) (court rejects appellants' contention that they met "the virtually impossible' burden of identifying alternative selection devices which satisfy the employer's legitimate hiring needs"). See also Rivera v. City of Wichita Falls, 665 F.2d 531, 538 (5th Cir. 1982).

case was brought against ETS and the Illinois Department of Insurance by five people who had failed the Illinois Insurance Licensing Exam and by the Golden Rule Insurance Company which claimed to have had trouble finding licensed minority agents to sell insurance in minority communities. Plaintiffs introduced preliminary evidence indicating that 78 percent of the white test-takers passed the life insurance test, in contrast to only 65 percent of black test-takers, and that 82 percent of the white test-takers passed the accident and health insurance test, compared to 55 percent of the black test-takers.55

The core of the Golden Rule settlement was a requirement that all items used on various administrations of the test be analyzed and classified into the following two categories:

1. Type I — those items for which (a) the correct-answer rates of black Examinees, white Examinees, and all Examinees are not lower than forty percent (40%) at the .05 level of statistical significance, and (b) the correct-answer rates of black Examinees and white Examinees differ by no more than fifteen (15) percentage points at the .05 level of statistical significance; or
2. Type II — all other items.56

After this classification had been accomplished, the decree then required ETS to assemble new test forms “in accordance with the subject matter coverage and weighting of the applicable content outline,” pursuant to the following guidelines:

1. Type I Items shall be used exclusively so long as they are available in sufficient numbers.
2. Those Type I Items for which the correct-answer rates of black Examinees and white Examinees differ least shall be used first.
3. Type II Items may be used, and shall be used before any new items . . . may be used, to the extent Type I Items are not available in sufficient numbers.
4. To the extent it is necessary to use Type II Items, those Type II Items for which the correct answer rates of black Examinees and white Examinees differ least shall be used first.57

In short, Golden Rule requires ETS to use questions on which blacks as a group tend to perform as well as whites before it uses items on which the performance differential is greater.

On its face, the settlement appears relatively innocuous. It permits Illinois to continue to use insurance licensing examinations. It

indirectly endorses ETS' validation methods and permits the company to assemble tests from its existing item banks. All that is specifically required is that items be drawn from the overall pool in a certain designated order.

Upon further analysis, however, the Golden Rule approach raises a number of serious concerns. The major problem involves the distortion of the proportionate weight of subject matter covered by an exam which is established by the content validation process to reflect "on-the-job" competence requirements. The weighting of the different subjects covered by the exam is known as "the blueprint." If many items necessary for the test blueprint are eliminated or modified, the integrity of this blueprint may be jeopardized. For example, if an analysis of a high school mathematics teacher's job indicates that 20 percent of her time will be spent teaching geometry, the test blueprint will require that 20 percent of the items on the test cover geometry. If blacks as a group fare relatively poorly on questions dealing with geometry, elimination of those items from the test will raise their scores. However, elimination of these items will also distort the validity of the test as an indicator of the competence of all teachers certified in mathematics to perform adequately on the job.\(^5\)

Test content may be further distorted by the preference of the Golden Rule approach for easy items. "Type I" items are more likely to be "easy" since questions that all groups answer correctly will not exhibit high differential statistics. Accordingly, difficult concepts will be measured less often, even if such concepts are an important measure of ability to perform on the job. Finally, the item bias approach may invite negative psychological and political reactions from white candidates who feel the technique is unfair because it eliminates questions on which they do well. This reaction is not unreasonable since current evidence indicates that the specific items eliminated under item bias techniques do not have any apparent culturally or racially biased wording or content.\(^5\)

\(^5\) See Cordes, supra note 55 at 26 (comments of Prof. Robert Linn raising blueprint issues).


These bills would result in severe adverse consequences for those individuals and educational institutions that objective tests are designed to serve. We have studied the most recent drafts of these pieces of legislation carefully. They are based on the erroneous assumption that differences in the proportions of students in various groups that answer test items correctly provide evidence that test items are biased
Teacher Competency Testing

In response to some of these concerns, ETS has attempted to minimize the significance of the Golden Rule settlement. An ETS spokesman concluded, "[T]here is nothing in the settlement requiring us to drop questions . . . The way we do business now has basically been confirmed." Plaintiffs, however, interpreted the settlement differently. They heralded it as a major breakthrough: "[I]t is a significant victory in the effort to eliminate discrimination in our country . . . [I]t is likely . . . that students taking the SAT and GRE will want similar safeguards, now that they know it can be done." Plaintiffs also noted the legal significance of the Golden Rule formula as "an alternative method" which would have to be considered by all developers of licensing or certification tests under the EEOC Guidelines. An expert psychologist retained by plaintiffs remarked that he expected the settlement to force ETS to revise its other tests because "once you have this method, to not use it is to knowingly use a more discriminatory test."

Widespread publicity about the settlement within the testing industry and a general awareness of the legal requirement to use alternative selection methods have created an item bias bandwagon. Test-makers are considering the Golden Rule approach when they develop new licensing examinations. Legislation to mandate the technique is pending in several states. Lawyers have begun to apply the

---

against members of these groups. These bills are also based on the assumption that items showing group performance differences of more than 15% and items that cannot be answered correctly by more than 40% of minority students, so-called "Type II" items, are "biased" against members of the lower scoring group and should be preferentially avoided in test construction. Both of these assumptions are seriously flawed. They do not correspond to measurement specialists' understanding of the meaning of group performance differences and do not lead to acceptable professional practice in test development and use.


60. Insurance Agent Tests Face Change to Settle Suit, Wall St. J., Nov. 29, 1984, at 3. ETS' news release on the day of the settlement explained its motivations for settling as follows:

"We are confident the examination would have withstood the scrutiny of the courts," said Stanford von Mayrhofer, general counsel of ETS, the non-profit organization which developed the test. "The lawsuit was dismissed twice by the courts in the eight years since the case began. The settlement found no fault with the test. We are not required to change the content of the examination but we may make such changes as we see fit. ETS will not pay any damages to the plaintiffs or any of the plaintiffs' costs" von Mayrhofer added. "In light of the favorable agreement reached, we saw no good reason to devote large amounts of time, energy, and money to continue the litigation."


62. Id.
concept in pending cases; for example, it was central to the recent Alabama teacher certification case consent decree.\textsuperscript{63}

The Alabama settlement, the validity of which is still being litigated,\textsuperscript{64} not only adopted the \textit{Golden Rule} method but substantially extended it. The settlement requires division of all items in the teacher competency test into three categories: Type I contains items with black/white performance differentials of no more than 5 percent; Type II contains items with ranges between 5-10 percent; and Type III consists of items ranging between 10-15 percent. In general, however, only the Type I and II items could be used in actual tests; inclusion of Type III items would be permitted only after the pool of Type I and II items had been exhausted, and, even then, no more than 10 percent of the total number of questions could be Type III. Items with a differential ratio above 15 percent could never be used.\textsuperscript{65}

In sum, the 15 percent differential analysis concept used in \textit{Golden Rule} to classify items for order of use was extended in the Alabama settlement to preclude use of any items falling outside that 15 percent range. Essentially, the Alabama approach precludes use of any items having a substantial adverse impact on minority candidates. Thus, the limited item bias approach of the Illinois case established a precedent that blossomed into a substantially more radical form in its very next application.

Given the political pressures to reduce the adverse impact of


\textsuperscript{64} This decree has had an unusual history and its final status is still unclear. Immediately after counsel for the State Education Department had agreed to the consent decree, the Attorney General objected both to its substance (among other things, it required the state to certify 500 members of the class who had failed the examination and to pay $500,000 in liquidated damages) and to counsel’s authority to bind the state to settle. Counsel’s consent was based on the position of the State Commissioner of Education, and on the lack of opposition from members of the State Board of Education when the proposed settlement had been discussed with them.

After the Attorney General publicized his objections, the members of the State Board of Education formally met and officially voted to reject the consent decree. Although initially ruling that their prior actions constituted an approval which was now binding, Allen v. Alabama State Bd. of Educ., 612 F. Supp. 1046, 1052 (M.D. Ala. 1985), the court, upon rehearing, held that in light of the broad public policy issues involved, it would vacate the settlement order and not impose the controversial consent decree on the now clearly unwilling defendants. The court also certified the questions for immediate appeal. Allen v. Alabama, Civ. No. 81-697 (Order of Feb. 4, 1986). That appeal is pending, as are the results of the trial which concluded in June 1986. (The author is represented NES, the test developer, as amicus curiae at the trial.) Both on appeal and at trial, the issues involving the item bias techniques set forth in the original settlement are likely to receive substantial attention.

\textsuperscript{65} Allen v. Alabama, Consent Decree of July 12, 1985, supra note 63, at ¶ 2(b)-(d).
Teacher Competency Testing

teacher certification exams, and the extensive advocacy campaign being mounted to adopt and extend the *Golden Rule* formula, it seems likely that the radical Alabama version rather than the limited Illinois approach will become the predominant model. Should the courts explicitly endorse the technique, especially the Alabama version, the item bias procedure might well be used as a valid alternative under the EEOC Guidelines, requiring its consideration and possibly its adoption in all teacher certification situations.

Such use of the item bias approach could compromise the basic purpose of teacher certification testing. The blueprinting distortions and the tendency to use easier items which is inherent in the technique would ensure certification of more applicants — both minority and non-minority — without assessing their competence by meaningful objective standards. Job access would be provided to more minority candidates, but at the cost of undermining certification standards and possibly lowering the quality of education of future generations of school children. Clearly, the public interest and ultimately the interest of the ill-prepared minority candidates themselves would be better served by pursuing reforms that would maintain the integrity of the tests.

IV. *Recommendations and Conclusion*

Although legal challenges to teacher competency tests on behalf of minority applicants appear to be proliferating, influential black leaders in the education community are speaking out against the anti-testing stance. Bernard R. Gifford, Dean of the Graduate School of Education at the University of California at Berkeley, for example, has argued that devaluing standardized tests evades the serious issues raised by high minority failure rates and will only per-

66. There is, however, a way in which the statistical techniques utilized in *Golden Rule* could be used to assist in minimizing any true bias without undermining the tests' basic integrity. Statistical indications of disparities in the performance of different racial groups (especially as reflected in the more sophisticated form of item bias analysis known as item response theory, which differentiates black and white test-takers of equal achievement levels rather than gross group statistics) can be used to "flag" items which need to be reconsidered by an expert review panel. *See generally Handbook of Methods for Detecting Test Bias* (R.A. Berk ed. 1982). Such a panel of knowledgeable persons (including a significant proportion of minority representatives) should carefully review the particular question to determine whether the cause of the statistical disparities is cultural bias in the wording of the question, or a deficiency in the educational preparation of the subject group of test-takers, or other factors. If cultural bias does exist, the item clearly should be reworded or eliminated. If biased phrasing or content cannot be articulated, however, the item should be retained so that the test's integrity is not compromised.
petuate the problem.\textsuperscript{67} Similarly, Mary H. Futrell, President of the National Education Association, an organization that had previously taken a public stand in opposition to teacher certification testing, recently stated:

I've heard some say that pre-service testing may hurt women and minorities. . . . As a black woman, I don't buy that. As a matter of fact, I resent it. If we set clear and demanding expectations and then help all potential teachers reach those expectations, we can have both quality and equality.\textsuperscript{68}

In short, these black leaders agree that, in the long run, maintenance of educational standards and of the integrity of certification tests will benefit both minority students and minority teachers.\textsuperscript{69}

Extending instructional validation or the \textit{Golden Rule} technique to teacher competency tests may, in the short run, raise minority passing rates, but it will ultimately undermine the reformers' efforts. Advocates for minority rights would be better advised to press a reform agenda to meet Gifford and Futrell's call for raising the substantive achievement level of minority applicants.\textsuperscript{70} Such reforms should have two major components. As a short-term program, states should provide minority students with specific test preparation assistance. In the long term, the basic educational opportunities afforded minorities must be improved.\textsuperscript{71}

\textsuperscript{67} Gifford, \textit{We Must Interrupt the Cycle of Minority Group Failure}, Education Week, Mar. 20, 1985, at 24.


\textsuperscript{69} Recent data on rising achievement test scores of minority applicants provide empirical support for the proposition that minority teacher candidates will fare better on the exams as their educational opportunities improve. For example, between 1976 and 1984, mean scores of black students on the Scholastic Aptitude Test (SAT) rose ten points on the verbal part of the test and nineteen on the mathematical, and mean score differences on the SAT between all students and black students were reduced by 13 percent on the verbal portion and 18 percent on the mathematical. Similarly, reading scores of black students tested at ages nine, thirteen, and seventeen improved steadily and consistently between 1970 and 1980. Anrig, \textit{Educational Standards, Testing, and Equity}, Phi Delta Kappan, May 1985, at 624.

\textsuperscript{70} Professor William Mehrens, past President of the National Council on Measurement in Education, has aptly stated the core issue as follows:

The solution to the problem of incompetent teachers be they black or white is to work at increasing their competence, not allowing them to teach in spite of their incompetence due to sympathy, guilt, or some perversion of the notion of justice. Mehrens, \textit{supra} note 13, at 70.

\textsuperscript{71} These recommendations are set forth as suggestions for further analysis. No claim is made that a definitive legal basis for enforcing them presently exists. The proposals are geared to the circumstances of entry level teacher candidates who are denied initial certification. The extent to which the situation of incumbent teachers, who in some states may be denied certification renewal if they fail a competency test, differs from that of entry level candidates is not discussed in this article.
Teacher Competency Testing

In all competency testing situations, extensive preparatory materials such as study guides, sample questions and counseling should be provided at state expense for minority students who have not had an opportunity to develop sophisticated test-taking skills. More important, an extensive program of remedial assistance aimed at improving basic skills in areas of identifiable weaknesses should be offered to all who fail an initial administration of the test. Where necessary, such a program might also provide funding for additional semesters of study to help committed candidates reach objective competency levels. Candidates should also be granted multiple opportunities to retake the test.

This meaningful approach to reform could also eliminate many of the problems posed by “instructional validation.” I have argued in this article that when there is a disparity between classroom job requirements and curriculum content at a teacher training institution, the job analysis requirements should prevail. There is no reason, however, why this disparity should exist. Institutions which purport to be dedicated to teacher training should gear their curriculum content to include, at a minimum, basic on-the-job competency requirements. State universities and teacher training colleges should be required to analyze their curricula and to gear curricular content to the job-related objectives of teacher certification examinations. Private institutions with curriculum offerings not substantially correlated with the job-related objectives established for state teacher certification examinations should be required to disclose this lack of correlation. Applicants for admission would then at least have notice that a degree from the institution would not be likely to provide them with thorough preparation for the state licensing examination.

Teacher candidates should also be given sufficient notice that a job-related teacher certification exam will be required. As in Debra P., where the court enjoined implementation of Florida’s student competency testing program for a four year period in order to provide fair notice and an opportunity to prepare for the examinations to minority students who earlier in their school career had been vic-

---

72. Specific diagnostic information concerning each student’s strengths and weaknesses should be provided to all candidates after administration of each test.

73. Such disclosure is required by canons of fairness and would not necessarily discourage all applicants from applying to the school, especially if it offered other academic benefits. For example, many law schools with strong national reputations do not purport to prepare their graduates for any particular bar examination; graduates of these schools generally assume that they will need to take an intensive short-term course on state law and procedures before sitting for a state bar examination.
tims of state-imposed segregation of the school system, fairness requires adequate notice. A similar due process concern was raised by the district court in United States v. Texas, when it noted that minority students had been given little notice of the Texas preprofessional skills test requirement before its effective date; that court's decision was, however, reversed by the Fifth Circuit.

Reasonable notice requirements should apply to all teacher competency tests. Basic equity considerations, if not constitutional due process requirements, call for special efforts to be made to maximize notice and opportunities for preparation when new standardized testing programs are initiated. The amount of notice considered reasonable will, of course, depend on the circumstances of the particular testing program. If the new program constitutes a radical departure from the past (for example, if newly promulgated competency standards were not previously an established part of the state teacher training college curriculum), a lead-in period equal to the length of the teacher training program would appear appropriate. In other situations (for example, when a testing program standardizes long-standing requirements) a one or two year notice period might be adequate.

Reasonable notice and enhanced test preparation should enable more minority candidates, especially those close to the cut score mark, to pass the tests. In the long term, however, the problem of the disparate impact of competency tests on minority candidates will only be solved by substantially upgrading the quality of education they receive. For graduates of the predominantly black teacher training institutions in the South, the failure rate on teacher competency tests is disproportionately high relative not only to white candidates but also to minority candidates attending integrated or predominantly white institutions. The reason for this disparity is


The district court also found that the State had denied the students due process by giving them little notice of, and no materials on, the nature of the tests, other than a pamphlet. While an individual is entitled to notice and a hearing before state action deprives him of life, liberty or property, no such right attends legislative enactments that affect a general class of persons. When the legislature enacts a law, or a state agency adopts a regulation, that affects a general class of persons, all of those persons have received procedural due process by the legislative process itself and they have no right to individual attention.

Lulac, slip op. at 7014.

76. See, e.g., Ayres, Student Achievement at Predominantly White and Predominantly Black Universities, 20 AM. EDUC. RESEARCH J. 291 (1983) (controlling for SAT scores, blacks attending predominantly white institutions averaged twenty-five points higher on NTE
simple: all of these institutions emerged from dual higher education systems. The legacy of inferior facilities and faculties which the predominantly black institutions inherited endures despite somewhat increased funding levels in recent years.\textsuperscript{77}

Some might argue that, given this bleak record, it would be better to close down the predominantly black institutions and steer minority applicants to predominantly white campuses.\textsuperscript{78} Despite the seeming logic of this approach, it should not be pursued. The predominantly black schools are the only avenue of opportunity available to thousands of minority students who, due to the inferior education provided to them in their early years, might not meet the entrance standards at more competitive institutions. Predominantly black colleges in the South must therefore be supported as they meet the needs of thousands of highly motivated minority students. The anticipated severe shortage of teachers, especially of minority teachers, projected for the next decade provides an additional rationale for such support. In order to provide real opportunity to the students at these colleges, the caliber of their faculty and the quality of their offerings must be enhanced and their funding levels substantially increased.\textsuperscript{79}

Improving predominantly black institutions in the South is only part of the long-term solution. In both the South and the North, affirmative action to assist minority students attending predominantly white teacher training institutions is also essential. Special scholarship programs that provide support and incentives to minority students who meet standards at competitive institutions must be expanded. To this end, Dean Gifford has proposed a plan which is currently under consideration by the California State Legislature. Gifford's plan would identify minority and low-income students who are committed to teaching and provide them with intensive university and post-graduate training. The plan also includes programs


\textsuperscript{78} Judge Clemon indicated that black students who attend many of the predominantly white institutions appear to be well integrated into campus life.

\textsuperscript{79} What is needed is a concerted campaign to attract larger numbers of academically-talented Black students into the teaching profession and to restructure the education departments at historically Black colleges. The competency testing and certification examinations that some consider 'necessary evils' may reveal some of the difficulties of minority students and help them better prepare for the future. Dilworth, Teachers Totter: A Report on Teacher Certification Issues, at 39 (Occasional paper of The Institute for the Study of Educational Policy No. 6, Howard University, 1984).
and rewards for outstanding teachers, non-minority as well as minority.\textsuperscript{80}

A final observation which must be made when considering any solution to the disparate impact of teacher certification tests on minority teacher candidates concerns the role of the courts in the development of professional standards. As they become immersed in institutional reform, courts are increasingly involved in determining complex social science controversies. Although empirical studies indicate that judges are able to obtain and comprehend complex social and factual data,\textsuperscript{81} the test validation and disparate impact issues raise new problems. Plaintiffs challenging teacher certification tests are asking judges not only to comprehend and apply established social science concepts, but to help formulate new professional standards.

Before Debra P. and Golden Rule, "instructional validity" and the "item bias analysis" were innovative suggestions which had been only tentatively raised in psychometric literature. They were not widely known, let alone widely accepted, before being thrust into prominence by the judicial process. Judicial imprimatur, even in a limited factual setting or in the confines of a settlement document, has conveyed an aura of legitimacy that substantially enhanced the professional acceptability of these concepts. Because of Debra P. and Golden Rule, these principles may become established fundamentals of professional judgment which must be carefully considered, if not adopted, by all practitioners in the psychometric field. The appearance of broad professional acceptance may well lead to further judicial invocation of these standards in future cases, on the assumption that they fairly reflect prevailing professional judgment in the field.\textsuperscript{82}

\textsuperscript{80} Gifford, supra note 67, at 24. \textit{Compare} Smith, supra note 7. Smith supports eliminating standardized certification tests and proposes affirmative action teacher recruitment programs.

\textsuperscript{81} See, e.g., M. REBELL AND A. BLOCK, EDUCATIONAL POLICY MAKING AND THE COURTS (1982).

\textsuperscript{82} The examples discussed in this article reflect a broader phenomenon that has marked the entire development of Title VII test validation doctrine. The original EEOC Guidelines incorporated evolving concepts of test validation proposed by a section of the American Psychological Association. The Supreme Court's decision in \textit{Griggs}, 401 U.S. 424, gave these procedures a supervening legitimacy. Within psychometric circles, the EEOC Guidelines became the incontrovertible core of professional standards. An analogous process has taken place in the area of I.Q. tests for student classification purposes. See Larry P. v. Riles, 495 F. Supp. 926 (N.D. Cal. 1979), aff'd, No. 80-4027 (9th Cir. Jan. 25, 1984); Prasse and Reschly, \textit{Larry P.: A Case of Segregation, Testing, or Program Efficacy?}, 52 EXCEPTIONAL CHILDREN 333 (1986); Bersoff, \textit{Regarding Psychologists Testify: Legal Regulation of Psychological Assessment in the Public Schools}, 39 MD. L. REV. 27 (1979); cf.
Teacher Competency Testing

Judges and professionals must recognize the interactive nature of this subtle process. Pronouncements of limited or qualified significance, based on professional understandings within one field, are overextended when too quickly translated into another. Given the dynamics of contemporary public policy formulation, judges can be important participants in the development of social science standards, and social scientists may have a legitimate interest in the development of legal doctrine. The interrelationship between social science and legal standards in the area of standardized competency tests should not be furthered, however, until all participants are fully aware of the nature of the dialogue and of the impact each discipline has on the other.