

1979

Competency Testing Programs: Legal and Educational Issues

Merle Steven McClung

Follow this and additional works at: <https://ir.lawnet.fordham.edu/flr>



Part of the [Law Commons](#)

Recommended Citation

Merle Steven McClung, *Competency Testing Programs: Legal and Educational Issues*, 47 Fordham L. Rev. 651 (1979).

Available at: <https://ir.lawnet.fordham.edu/flr/vol47/iss5/2>

This Article is brought to you for free and open access by FLASH: The Fordham Law Archive of Scholarship and History. It has been accepted for inclusion in Fordham Law Review by an authorized editor of FLASH: The Fordham Law Archive of Scholarship and History. For more information, please contact tmelnick@law.fordham.edu.

Competency Testing Programs: Legal and Educational Issues

Cover Page Footnote

Mr. McClung is currently working part-time as a staff attorney at the Center for Law and Education in Cambridge, Massachusetts and part-time as an education law consultant. The author acknowledges with appreciation the assistance of his colleagues at the Center, especially Robert Pressman, in developing the legal analysis that forms the basis of this Article, although this Article does not reflect their views in all parts. As of July 1979, Mr. McClung will be the director of the Law and Education Center, Education Commission of the States, Denver, Colorado.

COMPETENCY TESTING PROGRAMS: LEGAL AND EDUCATIONAL ISSUES*

MERLE STEVEN McCLUNG**

TABLE OF CONTENTS

INTRODUCTION	652
I. LEGISLATIVE, LEGAL, AND EDUCATIONAL BACKGROUND	653
A. <i>Legislative Background</i>	653
B. <i>Legal Background</i>	657
C. <i>Some Educational and Psychometric Considerations</i>	665
II. INAPPROPRIATE TEST CONTENT	672
A. <i>Coerced Belief</i>	674
B. <i>Invasion of Privacy</i>	677
C. <i>Unteachable or Unmeasurable Content</i>	678
III. ADEQUACY OF PHASE-IN PERIOD	679
A. <i>Issues</i>	679
B. <i>Analysis</i>	681
IV. MATCHING TEST AND INSTRUCTION	682
A. <i>Curricular and Instructional Validity</i>	682
B. <i>Basic Literary Skills Versus Adult Life Skills</i>	684
V. RACIAL AND LINGUISTIC DISCRIMINATION	687
A. <i>Racial Discrimination</i>	688
B. <i>Linguistic Discrimination</i>	692
C. <i>Cultural Bias</i>	694
VI. HANDICAPPED STUDENTS	698
A. <i>Exemption</i>	698
B. <i>Differential Diplomas and Standards</i>	699
C. <i>Differential Assessment Procedures</i>	699
D. <i>Other Legal Implications</i>	700
E. <i>Individualized Determinations</i>	701
VII. RECOMMENDED APPROACHES	701
A. <i>Community Participation</i>	702
B. <i>Two-Phase Implementation Procedure</i>	704
C. <i>Approaches to Instructional Validity and Assessment</i>	705
D. <i>Proposals for Competency Testing Programs</i>	708
1. <i>Basic Framework</i>	708
2. <i>Model Program Provisions</i>	709
CONCLUSION	711

* © Merle Steven McClung 1979.

** Mr. McClung is currently working part-time as a staff attorney at the Center for Law and Education in Cambridge, Massachusetts and part-time as an education law consultant. The author acknowledges with appreciation the assistance of his colleagues at the Center, especially Robert Pressman, in developing the legal analysis that forms the basis of this Article, although the Article does not reflect their views in all parts. As of July 1979, Mr. McClung will be the director of the Law and Education Center, Education Commission of the States, Denver, Colorado.

INTRODUCTION

COMPETENCY testing¹ is one of the most significant and controversial reform movements in public elementary and secondary education today. The movement is primarily initiated by state legislators who maintain that competency testing programs respond to public dissatisfaction with the large number of functionally illiterate students graduating from public high schools. By redirecting educational resources to students with poorly developed literacy and numeracy skills, some competency testing programs have constructive potential to improve student performance in essential basic skill areas. Many competency testing programs, however, have been designed and/or implemented in an inequitable manner and are likely to have more negative than positive effects. Programs that require a student to pass a minimal competency test as a prerequisite to a high school diploma, in particular, have potential for discrimination against students. Some of these programs may not only be unfair to students, they may be illegal as well.

Whether competency testing programs can and will have more beneficial than harmful effects remains an open question.² Some would agree with the National Academy of Education committee that competency testing as a prerequisite to a high school diploma "is basically unworkable, exceeds the present measurement arts of the teaching profession, and will create more social problems than it can conceivably solve."³

1. This Article uses the term "competency testing" in a general sense to mean tests purporting to measure basic skills and/or life-role activities because the legal analysis is essentially the same for both. More exact terminology is offered by William Spady who defines "*competencies* as indicators of successful performance in life-role activities (be they producer, consumer, political citizen, driver, family member, intimate friend, recreational participant, or life-long learner) and distinguishes them from the discrete cognitive, manual, and social *capacities* (such as reading and computational skills, speaking ability, and motivation) that, when integrated and adopted to particular social contexts, serve as the *enablers* or *building blocks* on which competencies ultimately depend." Spady, *Competency Based Education: A Bandwagon in Search of a Definition*, 6 Educ. Researcher 9, 10 (1977). The terms "proficiency" and "basic skills" are often used in the same sense as "capacities." While the legal analysis is essentially the same, the distinction between basic skills and life-role activities can be crucial. See notes 159-64 *infra* and accompanying text.

2. For a general introduction to the issues involved in minimum competency testing, see: American Friends Service Committee, *A Citizens' Introduction to Minimum Competency Programs for Students* (1978); National Institute of Education and Education Commission of the States, *Minimum Competency Testing: A Report of Four Regional Conferences* (1978) [hereinafter cited as *Minimum Competency Testing*]; Brickell, *Seven Key Notes on Minimum Competency Testing*, 59 Phi Delta Kappan 589 (1978).

3. National Academy of Education Committee on Testing and Basic Skills, *Report to the Assistant Secretary of Education: Improving Education Achievement 9* (1978) [hereinafter cited as *National Academy Report*]. This quotation refers specifically to "statewide minimum competency standards," but elsewhere in the report the same assertion is made with respect to competency testing generally. See, e.g., *id.* at 16.

Others would agree with the Educational Testing Service (ETS) that the problems are not insurmountable, and that a competency testing program establishing "a baseline below which relatively few students would fall, yet still allow students to acquire enough skills to function as citizens, is preferable to the status quo—an unacceptable quality of education."⁴

Whatever one's personal views, the movement is so strong that few elementary and secondary schools in this country are likely to be untouched by its impact—or by the concomitant legal and educational issues considered in this Article.⁵

I. LEGISLATIVE, LEGAL, AND EDUCATIONAL BACKGROUND

A. Legislative Background

A number of studies indicate that, whatever definition of literacy is used, substantial numbers of Americans are not literate.⁶ One of these studies, published by the Department of Health, Education and Welfare (HEW), concluded that an estimated one million American youths, twelve to seventeen years old, probably could not read as well as the average fourth grader, and thus they could be called illiterate.⁷ The study showed that disproportionate numbers of black youths were illiterate (15%), and that substantial numbers of white youths were also illiterate (3.2%).⁸ Not surprisingly, the study also found that the rate of illiteracy correlated with family income, declining "from 14

4. Educational Testing Service, Information Report: Basic Skills Assessment Around the Nation 3 (1977) [hereinafter cited as Basic Skills Assessment].

5. This Article is based in part upon the following earlier publications by the author: McClung, *Are Competency Testing Programs Fair? Legal?*, 59 Phi Delta Kappan 397 (1978); McClung, *Competency Testing: Potential for Discrimination*, 11 Clearinghouse Rev. 439 (1977) [hereinafter cited as *Potential for Discrimination*]; McClung, *Developing Proficiency Programs in California Public Schools: Some Legal Implications and a Suggested Implementation Procedure*, in Cal. State Dep't of Educ., Technical Assistance Guide for Proficiency Assessment app. K (1978); McClung & Pullin, *Competency Testing and Handicapped Students*, 11 Clearinghouse Rev. 922 (1978).

6. The exact degree of illiteracy, however, is subject to debate. Donald L. Fisher of the University of Michigan reports: "[H]igh schools have been accused of graduating thousands, even millions, of functional illiterates. . . . This accusation is, at first glance, supported by numerous surveys and observations. For example, the surveys referred to in this paper report that anywhere between 2 and 13 percent of the population with a twelfth grade education are functionally illiterate. We will argue that the upper estimate is greatly inflated, and more often than not misinterpreted." D. Fisher, *Functional Literacy and the Schools* 1 (1978). Of the studies that report racial data, however, almost all show that disproportionate numbers of blacks and Hispanics are illiterate. Another question that is subject to debate is whether schools are more or less successful today than in the past in teaching basic literacy and numeracy skills.

7. U.S. Dep't of Health, Education, and Welfare, *Literacy Among Youths 12-17 Years* 3 (1973).

8. *Id.* at 4.

percent in the lowest income group (less than \$3,000) to 0.3 percent in the highest (\$15,000 or more)."⁹

The widely publicized Adult Performance Level (APL) study conducted by the University of Texas at Austin¹⁰ found that on overall competency performance in five knowledge and four skill areas, 19.7% of the population could be classified as "functionally incompetent" or "adults who function with difficulty," 33.9% could be classified as "functional adults," and 46.3% could be classified as "proficient adults."¹¹ The functional incompetence rate was 21.7% in reading, 16.4% in writing, and 32.9% in computation.¹² The study concluded: "Overall, approximately one-fifth of U.S. adults are functioning with difficulty."¹³

Some studies exaggerate the extent of illiteracy and incompetency among minority groups. An example is the Texas study which noted great differences between whites and minority groups. "While 16% of the Whites are estimated to be functionally incompetent, about 44% of the Black and 56% of the Spanish-surnamed groups are estimated to be so. Here, as with other variables that have been discussed, the differences are probably due to the relatively lower levels of income, education, job status, and job opportunity found among minority groups in this country."¹⁴ This was only a partial explanation, however, because the extent of minority incompetence was exaggerated by questionable test norms, as discussed in Part V(C).

The HEW and Texas studies nevertheless identify serious shortcomings of many public schools in teaching and students in learning basic skills. There is no consensus about whether competency testing is the best means of correcting those shortcomings. Few would disagree, however, that functional illiteracy in today's world for large numbers of students after twelve years of public education is unacceptable. While higher level skills such as analysis, synthesis, and evaluation may be beyond the reach of many students, most educators would agree that basic cognitive skills and processes fall within the domain of teachable and learnable skills for virtually all students.¹⁵ Discussion concerning the broader goals of public education is suspect if basic

9. *Id.* at 6.

10. University of Texas at Austin, *Adult Functional Competency: A Summary (1975)* [hereinafter cited as *Texas Study*]. The Texas research was turned over to the American College Testing Program (ACT) of Iowa City, Iowa, a test developer and publisher. ACT publishes an Adult APL Survey and a Secondary APL Survey. ACT now markets tests based on the Texas research and also tests tailored to specifications provided by school districts.

11. *Id.* at 5-6.

12. *Id.* at 6.

13. *Id.*

14. *Id.* at 8.

15. See, e.g., Madaus & Airasian, *Issues in Evaluating Student Outcomes in Competency-Based Graduation Programs*, 10 *J. Research & Dev. Educ.* 79, 81 (Spring 1977).

literacy cannot be assured for virtually all students after twelve years of schooling.¹⁶

The various reports of widespread illiteracy have produced a dramatic reaction in the public and in state legislatures. The legislative rush to adopt competency testing programs has been startling. The Education Commission of the States (ECS) has issued a series of updates since April 1976 in an attempt to keep educators informed of developments in minimal competency testing. In an update published in November 1977, ECS reported:

As of April 15, 1976, four states had enacted legislation and four state boards of education had adopted regulations mandating some form of minimal competency testing. Bills were pending in eight state legislatures and five state boards had studies under way. Today, 11 states have enacted legislation and 20 state boards have adopted regulations, for a total of 31. Activity is occurring in the other 19 states either at the local level, by state department study or legislative hearings.¹⁷

There was also a proposed bill¹⁸ to amend the Elementary and Secondary Education Act of 1965¹⁹ to require all state agencies to establish a program of basic educational proficiency standards before funds could be received under the Act, but Congress limited the federal role to special financial assistance for states and local districts pursuing these goals.²⁰

The development of competency testing programs has occurred in three stages. Statewide assessment programs were first designed simply to measure performance in certain basic skills, for the purpose of identifying the school districts or individual schools in need of help. Extensive statewide programs of this kind were adopted in Colorado,²¹ Michigan,²² New Jersey,²³ and other states.²⁴

16. This statement is qualified to account for some mentally handicapped and other students who do not have normal capabilities. For a challenge to traditional assumptions of student capabilities, see B. Bloom, *Human Characteristics and School Learning* (1976). Bloom's research in mastery learning leads him to conclude that it is possible for 95% of all students to learn, at near the same mastery level, all that the school has to teach, with only 1 to 3% at the bottom level who cannot master the curriculum.

17. Education Commission of the States, *Update VII: Minimal Competency Testing 1* (C. Pihlo ed. 1977) [hereinafter cited as *Update VII*].

18. H.R. 6088, 95th Cong., 1st Sess., 123 Cong. Rec. H3093 (daily ed. Apr. 5, 1977) (introduced by Rep. Mottl).

19. Pub. L. No. 89-10, 79 Stat. 27 (current version at 20 U.S.C. §§ 236-246 (1976)).

20. See Education Amendments of 1978, Pub. L. No. 95-561, §§ 921-922, 92 Stat. 2296 (to be codified in 20 U.S.C. §§ 3331-3332). Section 921 authorizes federal assistance to the states in setting education proficiency standards, and § 922 authorizes federal assistance to state and local educational agencies to help develop their capacities to conduct programs for testing the achievement in basic skills of children in elementary and secondary schools.

21. Colo. Rev. Stat. §§ 22-7-101 to -105 (1974), discussed in Education Commission of the States, *Cooperative Accountability Project, Educational Accountability in Four States 3-7, 96-104* (1975) [hereinafter cited as *Educational Accountability*].

22. Mich. Comp. Laws Ann. §§ 388.1081-.1086 (1976), discussed in *Educational Accountability*, *supra* note 21, at 10-11, 108-09.

During the second stage in the development of competency testing programs, the burden of poor schooling has been shifted onto the student by testing each student as an individual, with the demonstration of minimal competency in basic skills being a prerequisite to a high school diploma and/or grade-to-grade promotion. A survey made in July 1978 reported thirteen states in this "second stage" of competency testing: Arizona (as of 1976), Oregon (1978), Florida (1979), North Carolina (1980), Utah (1980), Alabama (1981), California (1981), Delaware (1981), Vermont (1981), Virginia (1981), Maryland (1982), Nevada (1982), and Tennessee (1982).²⁵ And a February 1978 survey found: "Overall, the majority of the states, nearly 80 percent, are, or will be establishing in the near future, requirements for high school graduation based on competency examinations."²⁶ Two leading states in this movement are Florida²⁷ and California.²⁸

The two states nevertheless represent dramatically different approaches. Florida has statewide standards and a statewide test. In California the State Department of Education is precluded from developing a single statewide test—the responsibility for designing a test falls to each local district. This difference in approach has profound educational implications, as discussed in Part I(C). What is similar about the Florida and California mandates (requiring students to pass a competency test as a condition to receiving a regular high school diploma) are timelines that make it very difficult for districts to develop competency testing programs in a careful and sound way. The senior class of 1978-1979 will be the first to face the new Florida graduation requirements. In California the first class will be the one that graduates after June 1980.

23. N.J. Stat. Ann. §§ 18A:7A-4 to -16 (West Supp. 1978-1979), *discussed in* Education Commission of the States, Cooperative Accountability Project, *Legislation by the States: Accountability and Assessment in Education* 7, 13-22 (1975).

24. *See, e.g.*, Education Accountability Act of 1971, ch. 71-197, §§ 1-5, 1971 Fla. Laws 1148 (current version at Fla. Stat. Ann. § 229.55 (West 1977)), *discussed in* Educational Accountability, *supra* note 21, at 8-9.

25. *See* National Education Association, *Impact of Minimum Competency Testing in Florida*, *Today's Education*, Sept.-Oct. 1978, at 30, 33 [hereinafter cited as *Impact*].

26. Chall, *Minimum Competency Testing*, 22 *Harv. Graduate Sch. Educ. Bull.* 9, 10 (1978).

27. *See* Fla. Stat. Ann. §§ 229.55-58 (West 1977). The Florida program has been criticized on psychometric, legal, and educational grounds. *See* Glass, *Minimum Competence and Incompetence in Florida*, 59 *Phi Delta Kappan* 602 (1978) (psychometric critique); *Potential for Discrimination*, *supra* note 5 (legal critique); *Impact*, *supra* note 25 (educational critique); notes 63, 83-85, 172, 177, 183-88 *infra* and accompanying text. On October 16, 1978 a lawsuit challenging the Florida competency testing was filed. *Debra P. v. Turlington*, No. 78-892 Civ.-T-H (M.D. Fla., filed Oct. 16, 1978), *discussed at* note 63 *infra* and accompanying text. For a favorable view of Florida's program, see Fisher, *Florida's Approach to Competency Testing*, 59 *Phi Delta Kappan* 599, 599-602 (1978).

28. *See* Cal. Educ. Code §§ 51215, 51217 (West 1978 & Supp. 1979); Hart, *The California Pupil Proficiency Law as Viewed by Its Author*, 59 *Phi Delta Kappan* 592 (1978).

The third stage in the development appears to be a legislative compromise, satisfying the political need to respond in some way to the public demand for accountability, but avoiding the difficult decisions involved in how to use the test results. Whereas the second stage is characterized by the state mandating that passing a competency examination is a prerequisite to receiving a high school diploma, the third stage is characterized by the state mandating competency testing, but leaving the decision to tie the test to promotion or diploma to local districts. Massachusetts,²⁹ Connecticut,³⁰ Rhode Island,³¹ and New Hampshire³² are in the third stage of development. Regardless of whether the third stage approach is motivated mainly by a greater deference to local autonomy or by an increased sensitivity to the potential legal and educational problems, the result is the same for local districts. Local districts in third stage states will probably feel some pressure to tie the test results to promotion and/or diplomas. Therefore, the legal and educational issues considered in this Article may be helpful to local districts as well as to uncommitted states in deciding whether to tie prospective competency testing to promotion and/or diplomas, and, if they decide affirmatively, which approaches will minimize the concomitant problems.

B. *Legal Background*

In considering whether to develop some kind of competency testing program, one of the most crucial questions faced by a state or a local school district is how the test results will be used. The legality of a testing program will usually depend more upon the use of the test results than upon the test itself.³³ Using the test results as the primary basis for any decision that will cause serious harm to a student raises the initial legal questions. The trigger for legal analysis is this injury. Assuming there is injury, the following questions arise: who is responsible for that injury and does that person or agency have sufficient justification for causing that injury?

29. The Massachusetts State Board and Department of Education specifically state that their proposed regulations are *not* for the purpose of establishing a competency test as a new condition for promotion or graduation, but the proposed regulations do not prohibit use for this purpose. The purpose of the regulations is "to assist students in achieving mastery of basic skills prior to high school graduation through the provision of appropriate curriculum, instruction, and evaluation." Massachusetts Dep't of Education, Proposed Regulations: Basic Skills Improvement (1978) [hereinafter cited as Proposed Regulations].

30. An Act concerning Education Evaluation and Remedial Assistance, Pub. L. No. 78-194, 1978 Conn. Pub. Acts 258.

31. Rhode Island has not formalized its policy as of this writing.

32. See New Hampshire Dep't of Education, Guidelines for the Implementation of New Hampshire Accountability Plan (1978).

33. For a discussion of possible legal challenges to the test itself, see pts. II, V(C) *infra*.

If there is no injury, then there is no legal problem. Competency tests can be used in many ways that cause no injury to a student. For example, competency tests could be used simply to determine the general level of student performance in basic skills on a statewide or district level; to identify basic skill areas in an instructional program that need more emphasis; or to diagnose areas in which an individual student needs specific help. In such cases, there is usually no injury and no legal problem.³⁴

On the other hand, competency tests can be used to make decisions about individual students that have potential for grave injury. For example, competency tests can be used for tracking,³⁵ grade promotion, or denial of a regular high school diploma. Diploma denial, as mandated in Florida and California, probably causes the greatest injury to an individual student, and therefore raises the most serious legal questions.

Some people try to minimize the potential injury to students by saying that sending students into the world without basic skills is a greater disservice to them than diploma denial, and that a high school diploma based upon social promotion rather than achievement does not mean much anyway. This argument of course assumes that an effective educational program and academic achievement is dependent upon diploma denial,³⁶ and ignores the reality that many prospective employers and schools use the diploma as a minimum requirement without regard to the actual skills required for the task or possessed by the applicant.

Others will not try to minimize the injury of diploma denial, but will argue that schools have a responsibility to certify competence for employers, educational institutions and society at large, or that certification is a legally sufficient justification for a competency testing program.³⁷ Another justification likely to be advanced in support of

34. Tests with inappropriate content, however, may injure a student regardless of use. See pt. II *infra*.

35. For an extensive judicial analysis of the harmful effects of tracking, see *Hobson v. Hansen*, 269 F. Supp. 401 (D.D.C. 1967) (Wright, J.), *appeal dismissed*, 393 U.S. 801 (1968), *aff'd sub nom.* *Smuck v. Hobson*, 408 F.2d 175 (D.C. Cir. 1969). For other cases on tracking, see note 182 *infra*.

36. See note 246 *infra*.

37. Another common argument is that it is a misnomer to talk about "diploma denial" since a student has no legal right to a diploma, and therefore no legal basis for challenging denial of a diploma. A diploma, like education itself, however, is a benefit that must be made available to all on equal terms when the state decides to provide it. Consider the famous language of the Supreme Court in *Brown v. Board of Educ.*, 347 U.S. 483, 493 (1954): "[A]n opportunity of an education where the state has undertaken to provide it, is a right which must be made available to all on equal terms." Cf. *Griffin v. County School Bd.*, 377 U.S. 218 (1964) (unconstitutional closing of all public schools in county to avoid desegregation). Designating a diploma as a "privilege" rather than a "right" would not affect the constitutional analysis. See *Graham v. Richardson*, 403 U.S. 365, 375 (1971) (unconstitutional denial of welfare benefits to resident

competency testing is that it produces more beneficial than harmful effects. The argument is that a serious penalty is necessary to motivate students and teachers.³⁸ In this regard, Madaus and Airasian state: "If the minimal competencies are to be perceived as an integral aspect of the high school curriculum, skills to be consciously acquired and valued, the evaluation of competencies must possess two characteristics: it must be rigorous and it must involve penalties for non-mastery."³⁹ The strength of this argument will depend in part upon how successful the competency testing program is in helping students achieve the requisite skills.⁴⁰

In support of its claim that minimum competency testing programs can and do work, ETS points to two school systems—Westside High in Omaha, Nebraska and the Denver, Colorado public schools—as examples of successful programs.⁴¹ ETS states that Westside⁴² has graduated its first competency based class, with only 1% (eight seniors) of 762 in the senior class not passing; and that Denver, a pioneer in competency testing during the past eighteen years, has reduced its failure rate from 14% to 1.5%. The competency testing programs in Omaha (Westside) and Denver certainly merit careful study.

Whether other school systems will be able to similarly reduce their failure rate remains an open question. The high initial failure rates for students in Florida⁴³ and other geographic areas⁴⁴ where competency testing programs have been implemented are not encouraging, and the disproportionate failure rates for black and Hispanic students are alarming.

The use of competency tests to determine which persons are minimally competent and which persons are not, and to certify this

aliens). In sum, a state cannot withhold diplomas from some students without regard to federal constitutional safeguards.

38. The penalty, of course, falls on the student regardless of the quality of the teaching or the educational opportunities generally provided by the school. The disparities are most obvious between property-rich and property-poor school districts. To date, teachers have been sufficiently organized to prevent direct accountability for student performance on competency tests, but they are also concerned about being held indirectly accountable.

39. Madaus & Airasian, *supra* note 15, at 83.

40. Because individual rights are involved, a school presumably could not justify individual injury by demonstrating beneficial effects for students generally. The evaluation must ultimately be fair with respect to the individual denied a diploma. See notes 263-64, 267 *infra* and accompanying text.

41. Basic Skills Assessment, *supra* note 4, at 3. These and other competency testing programs are described in 59 Phi Delta Kappan 585-656 (1978).

42. Westside is located in a relatively affluent part of Omaha, and is not confronted with the difficult educational problems of most urban schools with large numbers of disadvantaged children.

43. See Florida Dep't of Education, *Statewide Assessment Results by Demographic Categories, Fall, 1977 Tests, Grade Eleven Highlights (Functional Literacy)* (1978).

44. Similar disproportionate failure rates on competency tests for minority students are reported by school systems in Virginia, North Carolina, and Milwaukee, Wisconsin.

competence or incompetence on a permanent record, could, in effect, create a permanent subclass of American citizens—a group of “second class” citizens comprised primarily of blacks, Hispanics, and other persons who do not enjoy the benefits of a white middle-class background. Some states will create three classes of citizens by offering a certificate of attendance as well: first class citizens who are deemed competent (and awarded a regular high school diploma), second class citizens who are deemed incompetent but dependable workers (certificate of attendance), and third class citizens who are simply labeled incompetent (no diploma or certificate). Some will argue that such distinctions are not created but simply identified and certified by the state. But this kind of competency testing program will, in effect, create many of the distinctions, and the state’s role in placing its official label of incompetence on a citizen cannot be trivialized,⁴⁵ especially because the state through its compulsory education laws, will penalize any underage student attempting to avoid the testing and labeling process.⁴⁶

The potential injury to students is great enough to raise sobering questions about whether to use competency testing for diploma denial or differentiation, even if the measurement technology were sufficiently developed to warrant confidence in determinations of competence and incompetence. The beneficial effects of a competency testing program should be considerable if it is to outweigh this probable injury. The injury is certainly serious enough to warrant special care in making determinations of incompetence. As underscored by a federal district court judge in another context: “When a program talks about labeling someone as a particular type and such a label could remain with him for the remainder of his life, the margin of error must be almost nil.”⁴⁷

For purposes of this Article, some form of serious injury to the student from the use of test results will be assumed. Denial of a high school diploma will be the primary example. Assuming that a plaintiff will be able to show injury from the use of the test results, another preliminary legal issue is whether the courts will refuse to review competency testing programs because they involve evaluations of academic performance.

Courts have traditionally been reluctant to review cases questioning the adequacy of educational programs or the validity of professional

45. The extent of the perceived injury of a certificate rather than a diploma is illustrated by comments attributed to Morris Milton, an attorney for the National Association for the Advancement of Colored People: “Kids would be better off dropping out of school at the end of their senior year than being stigmatized by a Certificate of Attendance as certified dummies.” Tallahassee Democrat, Mar. 24, 1978, § B at 3, col. 2 (quoting Mr. Milton).

46. The compulsory education laws in most states extend to age 16. See K. Alexander & K. Jordan, *Legal Aspects of Educational Choice: Compulsory Attendance and Student Assignment* 11-12 (1973).

47. *Merriken v. Cressman*, 364 F. Supp. 913, 920 (E.D. Pa. 1973).

judgments about academic performance of students. Such judicial restraint is rooted in an appreciation of the subjective nature of most determinations about educational quality and academic performance. The subjective nature of academic evaluations was a major factor in the failure of students to prevail in two recent suits against public school systems.

In *Peter W. v. San Francisco Unified School District*,⁴⁸ a functionally illiterate high school graduate failed in his \$1 million educational malpractice claim against the public high school that had presumably provided him with an inadequate education. Although the school had awarded him a high school diploma, he could not read and write well enough to fill out a standard job application. In affirming the dismissal of this case, the California Court of Appeals stressed the subjective nature of the relevant educational standards:

[T]he wrongs and injuries involved [are not] both *comprehensible and assessable* within the existing judicial framework. . . . Unlike the activity of the highway or the marketplace, classroom methodology affords no readily *acceptable standards of care, or cause, or injury*. The science of pedagogy itself is fraught with *different and conflicting theories* of how or what a child should be taught, and any layman might—and commonly does—have his own emphatic views on the subject.⁴⁹

48. 60 Cal. App.3d 814, 131 Cal. Rptr. 854 (1976).

49. *Id.* at 824-25, 131 Cal. Rptr. at 860-61 (emphasis added) (citations omitted). This language has important implications for *Peter W.*-type cases that may be brought against school systems that are implementing competency testing programs. The lack of objective standards in education is one reason why no educational malpractice case of this kind has ever been successful. See *Donohue v. Copiague Union Free School Dist.*, 64 A.D.2d 29, 33, 407 N.Y.S.2d 874, 877 (1978) (citing *Peter W.* in affirming dismissal of case).

Competency testing programs significantly change the nature of educational standards by setting explicit educational goals, clearly specified performance objectives related to those goals, objective test specifications to measure those objectives, and test items keyed to those specifications. The Florida statute arguably creates an explicit right to a minimally adequate education: "The first priority of the public schools of Florida shall be to assure that all Floridians, to the extent their individual physical, mental, and emotional capacities permit, shall achieve mastery of the basic skills. The term 'basic skills,' for purposes of this section, means reading, writing, and arithmetic." Fla. Stat. Ann. § 230.2311(1) (West 1977). Even in the absence of such statutory language, the officially stated purpose of most competency testing programs is to provide a minimum level of education for all students, and objective standards are adopted to serve that purpose.

Educators often express the concern that they may be sued whether or not they give a functionally illiterate student a high school diploma. Competency testing programs, with their explicit goals and objective standards, will provide a stronger basis for this concern. In special education, where professional standards are often more established, a student recently won an education malpractice case. A New York City jury awarded \$750,000 in damages to a former student who claimed that he had been negligently misclassified and miseducated as retarded. The New York Appellate Division reduced the award to \$500,000 but distinguished the *Donohue* and *Peter W.* cases and let the verdict stand. *Hoffman v. Board of Educ.*, — A.D.2d —, —, 410 N.Y.S.2d 99, 110-11 (1978). See generally Abel, *Can a Student Sue the Schools for Educational Malpractice?*, 44 Harv. Educ. Rev. 416 (1974); Ratner, *Remedying Failure To Teach Basic Skills*, 17 Inequality Educ. 15 (June 1974); Sugarman, *Accountability Through the Courts*, 82 Sch. Rev. 233 (1974); Comment, *Educational Malpractice*, 124 U. Pa. L. Rev. 755 (1976).

A similar rationale underlies the Supreme Court's refusal in *Board of Curators of the University of Missouri v. Horowitz*⁵⁰ to provide substantial procedural safeguards to medical students dismissed for academic reasons. Writing for a divided Court,⁵¹ Justice Rehnquist emphasized "the significant difference between the failure of a student to meet academic standards and the violation by a student of valid rules of conduct."⁵² Academic evaluations of a student, in contrast to disciplinary determinations, he noted, do not lend themselves to judicial factfinding procedures:

The decision to dismiss [the plaintiff] . . . rested on the academic judgment of school officials that she did not have the necessary clinical ability to perform adequately as a medical doctor and was making insufficient progress toward that goal. *Such a judgment is by its nature more subjective and evaluative than the typical factual questions presented in the average disciplinary decision.* Like the decision of an individual professor as to the proper grade for a student in his course, the determination whether to dismiss a student for academic reasons requires an *expert evaluation of cumulative information* and is not readily adapted to the procedural tools of judicial or administrative decisionmaking.⁵³

Relying on cases like *Peter W.* and *Horowitz*, some educators question whether competency testing is different from any other testing program traditionally implemented by public schools free of judicial intervention. In addition to claims of racial discrimination,⁵⁴ there are at least three important distinctions: the extent of the injury, the ultimate reliance on a single assessment, and the purportedly objective nature of that assessment. What makes competency testing different from most other high school testing programs is the severe nature of the penalty for failing a single test—denial of a high school diploma or its equivalent.⁵⁵ Until recently, very few public high schools in this country have made a student's diploma depend upon the outcome of one test,⁵⁶ and the severe nature of the penalty almost guarantees that the testing program will be subject to intense scrutiny by students, parents, and their lawyers.

50. 435 U.S. 78 (1978).

51. Justices Powell, White, Marshall, and Blackmun filed separate opinions. *Id.* at 79.

52. *Id.* at 86.

53. *Id.* at 89-90 (emphasis added).

54. *See* pt. V *infra*.

55. *See* notes 33-47 *supra* and accompanying text.

56. Some people point to the traditional Regents Examination in New York State, *see* N.Y. Educ. Law § 209 (McKinney 1969), as being similar to contemporary competency testing programs. The Regents Examination, however, is not used as a basis for denying high school diplomas or for determining whether a student is competent or incompetent. A Regents diploma is generally regarded as a superior diploma and has special value for college admissions. In May 1976, the New York Board of Regents approved a resolution making passage of a competency test a prerequisite to a high school diploma beginning with the graduating class of June 1979. *See* Update VII, *supra* note 17, at 39.

The most important distinction between contemporary competency testing and traditional testing programs, however, is the purportedly objective nature of the former. Most competency testing programs are specifically designed to provide relatively objective criteria for making the crucial evaluation of academic performance. Unlike the subjective evaluations involved in the *Peter W.* and *Horowitz* cases, competency testing programs purport to provide relatively objective measurements of competence. They are usually characterized by clearly specified standards: explicit educational goals, performance objectives related to those goals, test specifications designed to measure those objectives, test items keyed to those specifications, and finally the crucial determination of a cut-off score separating those deemed to be competent, from those deemed to be incompetent. These efforts are supported by psychometric specialists who perform the various psychological, statistical, and validation studies necessary to construct a test instrument that meets accepted professional standards. The result is a test instrument that appears as impressive and scientific as the best of the traditional standardized tests⁵⁷ but is also subject to many of the same as well as some additional shortcomings. Reliable, valid, and timely measurement of functional competency is especially problematic, as noted by William Spady of the National Institute of Education: "The technology surrounding the assessment and measurement of success in life role activities is only in its infancy, even though the rush toward adopting [competency-based] programs is upon us."⁵⁸

Although the technology of constructing competency tests is not sufficiently developed to warrant great confidence in its scientific objectivity, educators implementing many competency testing programs will be making assumptions of objectivity, and in turn making crucial decisions based on those assumptions. Those assumptions are likely to be tested in court. As a result, courts will no longer be faced with a challenge to a graduation decision based upon "an expert evaluation of cumulative information"—that is, hundreds of tests in scores of classes graded by numerous teachers using complex personal mixes of objective and subjective criteria. Instead, courts will be asked to review an obviously crucial decision based primarily upon the results of one test instrument purported to incorporate test items that are objectively related to clearly specified performance objectives and explicit educational goals. The academic evaluation involved in a competency test, therefore, seems, to use Justice Rehnquist's language, readily adaptable "to the tools of judicial decisionmaking."⁵⁹ To so

57. "The average American parent has a great and naive faith in 'scientifically' constructed tests." Sherrer & Roston, *Some Legal and Psychological Concerns About Personality Testing in the Public Schools*, 30 Fed. B.J. 111, 114 (1971).

58. Spady, *supra* note 1, at 11.

59. 435 U.S. at 90.

characterize competency testing is not to predict that the courts will in fact treat academic decisions based upon competency test results differently from traditional evaluations of academic performance; it does suggest, however, that the courts' primary rationale for past restraint is far less persuasive in the present situation, and that judicial review is warranted.

Apart from these new characteristics of competency testing programs, they will be subject to review by the legal standards applied to traditional testing programs. The most obvious of these standards are those prohibiting racial and linguistic discrimination in testing programs—in particular the constitutional prohibition against certain practices that carry forward the effects of proven, past discrimination.⁶⁰ In addition, procedural and substantive due process standards may provide a basis for scrutinizing competency testing programs.⁶¹

The minimal competency requirement as a prerequisite to a high school diploma is a new phenomenon in most states. Therefore, with the exception of racial and linguistic discrimination, it is difficult to identify the strongest legal arguments for or against it, and even more difficult to predict the judicial response. The first direct legal challenge to a state competency testing program,⁶² *Debra P. v. Turlington*,⁶³ was

60. See notes 179-88 *infra* and accompanying text.

61. See pts. III, IV *infra*.

62. Two limited procedural challenges to the Florida competency testing program preceded *Debra P. v. Turlington*, No. 78-892 Civ.-T-H (M.D. Fla., filed Oct. 16, 1978); *Robinson v. School Bd.*, No. 78-2137-CA (L) 01 H (Fla. Cir. Ct., filed June 8, 1978); *Brady v. Florida State Bd. of Educ.*, No. 78-653R (Fla. Div. Adm. Hearings: June 15, 1978).

In *Robinson*, plaintiff alleges that the Palm Beach County School Board did not provide adequate notice of the rule-making involved in requiring students to pass a competency test by June 1978 as a prerequisite to a high school diploma, and that the Board did not prepare an economic impact statement for the proposed rule as required by the Florida Administrative Procedure Act, Fla. Stat. Ann. § 120.54 (West Supp. 1974-1978).

In *Brady*, the court ruled that the cut-off score required by the board of education for evaluating mastery of the competency testing program is a "rule" within the meaning of Fla. Stat. Ann. § 120-52(14) (West Supp. 1974-1978). Because the scoring criterion was not promulgated or adopted in accordance with the Administrative Procedure Act, the court declared it an invalid rule, thus forcing the state to hold hearings on the rule pursuant to the Act. The state held hearings on the proposed rule in September 1978.

63. No. 78-892 Civ.-T-H (M.D. Fla., filed Oct. 16, 1978). The first sentence of the complaint's preliminary statement reads: "This is a class action brought by certain black school children in Hillsborough County challenging the defendants' administration of the Florida Functional Literacy Examination (now denominated as Florida Student Assessment Test, Part II), and the statutory scheme under which the test is required, as racially discriminatory and a deprivation of due process." Complaint at 1-2, *Debra P. v. Turlington*, No. 78-892 Civ.-T-H (M.D. Fla., filed Oct. 16, 1978) [hereinafter cited as Complaint]. The ten named plaintiffs seek to represent "a class consisting of all present and future twelfth grade black public school students in the State of Florida who have failed or who hereafter fail the Florida Functional Literacy Examination." *Id.* at 3. The named defendants include Commissioner of Education Ralph Turlington, the Florida State Board of Education and Department of Education, and the School Board of the Hillsborough County Schools. Plaintiffs ask the court, *inter alia*, to enter a

recently filed in federal district court in Florida. The plaintiffs allege that the Florida competency testing program as applied to black students violates rights guaranteed them by federal statutory and constitutional law. *Debra P.* is only the first in what is likely to be a long line of legal challenges as various states begin to withhold high school diplomas on the basis of a competency test, and a judicial pattern and predictable case law on competency testing will eventually emerge as a result of these challenges.

C. *Some Educational and Psychometric Considerations*

As of this writing, there are no court decisions or HEW guidelines setting minimum standards for competency testing of high school students, but both are likely in the near future.⁶⁴ Until then, decisions and standards in related areas offer guidance. Judicial involvement in determinations of test validity has focused primarily on situations involving employee selection procedures. The Supreme Court⁶⁵ has relied upon the Guidelines on Employment Selection Procedures (Federal Guidelines)⁶⁶ developed by the Equal Employment Opportunity Commission, and upon the Standards for Educational and Psychological Tests (APA Standards),⁶⁷ developed by a joint committee of the American Psychological Association (APA), the American Educational Research Association (AERA), and the National Council for Measurement in Education (NCME). In fact, the Federal Guidelines defer to the APA Standards on technical matters, as noted with some concern by Novick:

[T]he *Federal Executive Agency Guidelines* on employment selection issued in 1976 refer to the joint test standards and, in fact, defer to these standards for technical detail. Thus, executive agency guidelines effectively having the force of law are adopting professional standards that were not specifically written with this purpose in mind.⁶⁸

permanent injunction restraining defendants from requiring that members of the plaintiff class pass the examination as a prerequisite to receiving a high school diploma. *Id.* at 18. The complaint's specific claims for relief are set forth in text accompanying notes 145, 172, 183-88, 251 *infra*.

64. HEW has circulated a draft policy on competency testing to state departments of education for comment, but no policy has been approved as of this writing.

65. *See, e.g.*, *Washington v. Davis*, 426 U.S. 229 (1976) (American Psychological Association's (APA) Standards for Educational and Psychological Tests applied to testing practices of the Washington, D.C. Police Department in selecting individuals for its training academy).

66. 29 C.F.R. §§ 1607.1-14(1978).

67. American Psychological Association, *Standards for Educational and Psychological Tests* (1974) [hereinafter cited as APA Standards].

68. Novick, *The Influence of the Law on Professional Measurement Standards*, in *Educational Measurement & the Law* 41, 47 (1977).

Recent decisions by the Supreme Court indicate an increasing awareness of important differences between the Federal Guidelines and APA Standards, and also a split within the Court as to which set of test standards should be accorded greater deference.⁶⁹ This difference concerning the respective weight that should be given to federal agency guidelines versus professional standards is unlikely to be resolved soon, especially because the Federal Guidelines have been replaced by the new Uniform Guidelines on Employee Selection Procedure⁷⁰ and because the APA Standards are being considered for revision in light of their increased use in resolving legal disputes.⁷¹ Whatever the ultimate resolution, it is likely to incorporate considerable deference to professional standards on technical, psychometric questions. Therefore, some understanding of the basic psychometric concepts and terminology that courts have and will probably continue to rely upon in deciding questions of fair testing is essential.

The APA Standards define the requirements for reliable and valid tests. Simply put, *reliability* refers to whether the instrument measures accurately what it measures; for example, the instrument should yield comparable results when used at different times.⁷² *Validity* refers to whether the instrument measures what it purports to measure. There are many different kinds of test validity, as indicated below, and each should be considered in relation to the exact test in question. According to the APA Standards:

Questions of validity are questions of what may properly be inferred from a test score The many types of validity questions can, for convenience, be reduced to two: [w]hat can be inferred about what is being measured by the test [and] [w]hat can be inferred about other behavior?

The kinds of validity depend upon the kinds of inferences one might wish to draw from test scores. Four interdependent kinds of inferential interpretation are traditionally described to summarize most test use: the *criterion-related* validities (*predictive* and *concurrent*); *content* validity; and *construct* validity.⁷³

Predictive validity is a measure of how well test items predict the future performance of test takers. This type of assessment requires an analysis comparing the predictions about each test taker based on the test results with the actual functioning of the test taker at a later point in time. For example, if a high number of students which an Adult

69. See Lerner, *The Supreme Court and the APA, AERA, NCMA Test Standards: Past References and Future Possibilities*, 33 *American Psychologist* 915 (1978).

70. 43 Fed. Reg. 38290-309 (1978) (to be codified in 5 C.F.R. § 300.103(c); 28 C.F.R. § 50.14; 29 C.F.R. § 1607; 41 C.F.R. § 60-3).

71. An AERA, APA, NCME Review Committee has been established to make recommendations about the proposed revision of the APA Standards. See Letter from Willo P. White, Administrative Associate, Office of Scientific Affairs, APA to Merle McClung (Feb. 23, 1978).

72. See APA Standards, *supra* note 67, at 48.

73. *Id.* at 25-26 (footnote omitted).

Performance Level (APL) test had predicted would be functionally incompetent in fact turned out to be such when studied years later, then the test could be said to have predictive validity.⁷⁴

Concurrent validity is a measure of how well test results correlate with other criteria which might provide the same type of information about test takers. This type of assessment provides a measure of a test's immediate predictive validity, or how well determinations based upon test results correlate with other currently available information about test takers.⁷⁵

Construct validity is a measure of how well test items correlate to the theory or constructs behind the test. This assessment indicates the relationship between the test and actual test performance. This assessment is probably the most difficult to conduct since it may be difficult to identify the constructs upon which a test is built and because a statistical analysis of the interrelationship of test items may be required.⁷⁶

Content validity is a measure of how well test items represent the knowledge that the test purports to measure. A test with a high degree of content validity is a test for which high test performance serves as an index of a high degree of skills or knowledge in the area which the test purports to measure.⁷⁷

Two kinds of content validity not specified in the APA Standards—curricular and instructional validity—are of special relevance to competency testing. As argued below in Part IV, any competency test that lacks instructional validity in particular should not be used as the basis for a decision as important as diploma denial.

Two other issues with far-reaching educational and psychometric implications are whether a state committed to competency testing chooses (1) statewide or locally developed tests, and (2) criterion-referenced or norm-referenced tests. The issue of statewide versus local tests, with corresponding implications for statewide versus local curriculum, raises, in a new context, the classic controversy between state control and local autonomy, with local districts ever ready to ward off any attempted encroachment upon their traditional authority.⁷⁸

One advantage of a statewide test as in Florida is uniformity. Students moving from one school district to another within the state

74. See generally *id.* at 26. For comment on predictive validity and competency tests, see P. Airasian, G. Madaus, & J. Pedulla, *Policy Issues in Minimal Competency Testing and a Comparison of Implementation Models 73-74* (1978) [hereinafter cited as *Policy Issues*] (prepared for and published by the Massachusetts Department of Education).

75. See generally APA Standards, *supra* note 67, at 26.

76. *Id.* at 29-30.

77. *Id.* at 28-29.

78. On this issue, however, some local districts would prefer a statewide test because of the difficulty and expense of constructing a valid test. In fact, a number of local school districts in Virginia reportedly expressed this view to the state legislature and Department of Education.

will not be exposed to easier,⁷⁹ more difficult, or simply different standards, and thus employers and others throughout the state will know what minimum standards are certified by the diploma. One disadvantage is that a statewide test cannot be tailored to the diverse instruction and curricula that has been offered in local districts, and it will take a number of years for curriculum and instruction to adapt themselves to the test.

This in turn raises another classic controversy about whether tests should lead or follow curriculum, and whether teachers should "teach to the test."⁸⁰ Some commentators like Herbert Rudman would not see these as important issues with minimal competency testing because they do not see significant differences among American schools in basic skills curriculum and instruction. Others like Professor W. James Popham of the University of California at Los Angeles challenge this view:

At a rather general level of abstraction, Rudman is accurate in his claim that "there is a relatively uniform program among U.S. school districts." But day-to-day classroom activities are not fashioned at such a general level, nor are standardized test items written at this level of generality. It is only when we take such general concepts as "the ability to perform fundamental mathematical operations" and write test items or design instructional materials that differences in educators' preferences begin to show up. And there are differences aplenty.⁸¹

Differences between local districts in basic skill instruction, and between the test and instruction, will be especially acute where the competency test is designed to measure adult-life skills, as discussed in more detail in Part IV(B).

Thus statewide tests, at least in the short run, are inherently vulnerable to a charge of insufficient match between what the test measures and what the students have been taught. Locally developed tests, on the other hand, can be tailored to local curriculum and instruction, but designing a reliable and valid test that meets professional psychometric standards is beyond the expertise and perhaps the budget of most local districts. Test publishers are attempting to solve this problem by creating "item pools," as discussed herein.⁸²

Many educators, however, question whether current educational technology is sophisticated enough to produce a sound competency testing program, even if statewide resources are available. Others are engaged in an intense debate over the respective advantages and disadvantages of criterion-referenced versus norm-referenced tests—the former are usually preferred for competency testing programs.

79. Some reports indicate student flight to schools with lower standards. See *Time*, Dec. 12, 1977, at 22 (Florida juniors are fleeing to high schools in Georgia where a competency test is not required for graduation).

80. See generally Madaus & Airasian, *supra* note 15, at 82-90.

81. Popham, *The Standardized Test Flap Flop*, 59 *Phi Delta Kappan* 470, 471 (1978).

82. See notes 261-62 *infra* and accompanying text.

The intensity of the debate is evident in a hard-hitting article by Professor Gene Glass of the University of Colorado.⁸³ Glass criticizes the ETS item writers for arbitrarily writing relatively difficult mathematical items for the Florida test. Glass's main objection is the arbitrariness inherent in setting standards for criterion-referenced tests. "Teachers and their consultants attempting to define 'competencies' and writing test items intended to reflect minimal levels of acquisition are engaged in a bootless and potentially embarrassing endeavor."⁸⁴ Referring generally to Florida's competency testing program, Glass concludes that Floridians should call for "an immediate suspension of the minimal competence graduation requirements program, because it is based on indefensible technology. The items of the test have never been validated as measures of 'survival skills,' and the pass-fail standards were set mindlessly and capriciously."⁸⁵

Glass's claim of the indefensible arbitrariness of criterion-referenced tests is challenged by many,⁸⁶ including Professor Popham:

The cornerstone of Glass's attack on the setting of performance standards is his assertion that such standards are set *arbitrarily*. He uses that term in its most pejorative sense, that is, equating it with mindless and capricious action. But while it can be conceded that performance standards must be set *judgmentally*, it is patently incorrect to equate human judgment with arbitrariness in this negative sense.⁸⁷

Reviewing this debate, Professor Robert Linn of the University of Illinois concludes: "The process of setting standards and of making inferences based on those standards lacks a firm theoretical foundation, and this in turn creates problems that are apt to be of more than academic interest. . . . Typical current practice is not likely to provide a basis that most standard setters would be comfortable defending in court."⁸⁸

While Glass contends that criterion-referenced tests like the Florida test are "capricious and authoritarian,"⁸⁹ value judgments based on

83. Glass, *supra* note 27, at 602-05.

84. *Id.* at 603. For more detailed discussion of the state of the "science" of setting cut-off scores, see: Burton, *Societal Standards*, 15 J. Educ. Measurement 263 (1978); Glass, *supra* note 27. But see Block, *Standards and Criteria: A Response*, 15 J. Educ. Measurement 291 (1978); Hambleton, *On the Use of Cut-Off Scores with Criterion-Referenced Tests in Instructional Settings*, 15 J. Educ. Measurement 277 (1978); Popham, *As Always, Provocative*, 15 J. Educ. Measurement (1978) [hereinafter cited as *As Always*]; Scriven, *How to Anchor Standards*, 15 J. Educ. Measurement 273 (1978). For an article that raises some broader questions about standard setting, see Levin, *Educational Performance Standards: Image or Substance?*, 15 J. Educ. Measurement 309 (1978).

85. Glass, *supra* note 27, at 605.

86. See Block, *supra* note 84; Hambleton, *supra* note 84; *As Always*, *supra* note 84; Scriven, *supra* note 84.

87. *As Always*, *supra* note 84, at 298.

88. Linn, *Demands, Cautions, and Suggestions for Setting Standards*, 15 J. Educ. Measurement 306 (1978).

89. Glass, *Standards and Criteria*, 15 J. Educ. Measurement 237, 259 (1978).

comparative evidence as exemplified by norm-referenced tests impress him as "cogent and fair."⁹⁰ This assertion is challenged by critics of norm-referenced measures who point out, among other things, that norm-referenced tests raise similar difficulties and are less useful than criterion-referenced measures to diagnose learning problems and guide instruction.⁹¹ Critics also note that in order to produce a bell-shaped curve, norm-referenced tests are usually constructed to eliminate items that most students can answer and, therefore, are designed to exaggerate differences in achievement among students.⁹² Criterion-referenced tests, they argue, overcome such shortcomings by measuring students against objective standards rather than against themselves (a driver's license test being the classic example).⁹³

The relative merits and shortcomings of criterion-referenced versus norm-referenced measures was the subject of the 1978 AERA Presidential Debate. Professor Robert Ebel of Michigan State University argued that norm-referenced tests enjoy an established and superior technology.⁹⁴ Professor Popham, on the other hand, argued that criterion-referenced tests are superior for purposes of individual diagnosis, classroom instruction, and program evaluation.⁹⁵

Whether criterion-referenced or norm-referenced, most authorities recognize the subjective nature of setting cut-off scores for competency tests:

[A]lthough at first glance it appears as if the statistical technology is available for determining cut-scores, there is no consensus on any one approach and all approaches rely on subjective judgments of some sort. . . . In reality, part of what enters into the determination of a cut-score is the answer to the question "What percentage of students can we reasonably fail?"⁹⁶

The largely subjective and political questions of what competencies to test for and what is an appropriate cut-off score are two of the most important decisions in designing any competency testing program.

90. *Id.*

91. See Block, *supra* note 84; Hambleton, *supra* note 84; *As Always*, *supra* note 84.

92. See generally Buros, *Fifty Years in Testing: Some Reminiscences, Criticisms, and Suggestions*, 6 *Educ. Researcher* 9, 12 (1977) ("These statistical methods of item validation confuse differentiation with measurement and exaggerate differences among individuals and between grades. . . . In a 1948 paper . . . I urged the abandonment of these statistical methods of validating test items as well as the practice of discarding items simply because they were either passed by all or failed by all.").

93. See, e.g., W. Popham, *Criterion-Referenced Measurement* (1978).

94. Professor Ebel reiterates his belief that "a test of minimum competency should be a domain-sampling test, not an objective-referenced test. It should be designed to indicate a pupil's general level of achievement, not the particular competencies the pupil may or may not have learned." Ebel, *The Case for Minimum Competency Testing*, 59 *Phi Delta Kappan* 546, 548 (1978).

95. See generally W. Popham, *supra* note 93.

96. Policy Issues, *supra* note 74, at 60-61; Haney & Madaus, *Making Sense of the Competency Testing Movement*, 48 *Harv. Educ. Rev.* 462, 466-69 (1978).

Various individuals and groups within a given community (e.g., minority groups, employers) may have different views on these matters, and their support may be essential for an effective program. The inherently subjective and political nature of setting standards for competency tests provides a compelling argument for representative community-based participation in the process.⁹⁷

Some educators would not see much relevance in the debate over criterion-referenced versus norm-referenced tests because they do not believe that the measurement technology is sufficiently sophisticated to design any competency testing program sound enough to make crucial decisions such as which students receive and which are denied a high school diploma.⁹⁸ This view is articulated in a report by the National Academy of Education Committee on Testing and Basic Skills that was prepared at the request of HEW Secretary Joseph Califano.⁹⁹ The committee,¹⁰⁰ chaired by Stephen K. Bailey of the Harvard Graduate School of Education, concluded in part:

[A]ny setting of state-wide minimum competency standards for awarding the high school diploma—however understandable the public clamor which has produced the current movement and expectation—is basically unworkable, exceeds the present measurement arts of the teaching profession, and will create more social problems than it can conceivably solve.

. . . . The effort to determine and assure minimum competency standards for high school graduation will . . . fall of its own weight, for the scaffolding of existing test designs is too weak to carry such an emotionally-laden and ambiguous burden. Continuing extensive efforts and funds in this direction is wasteful and takes attention away from the major tasks of improving our schools.¹⁰¹

Reports such as these have dissuaded some states and school districts from following the Florida and California examples of using competency tests as a prerequisite to granting a high school diploma. But others continue to express an interest in the possible merits of such competency testing programs. Their interest is encouraged by organizations such as ETS, which helped to design the Florida test.

ETS takes the position that the difficulties raised by competency

97. See pt. VII(A) *infra*.

98. Consider, for example, the following statement by McKenna: "As promising as the concept of criterion-referencing seems for correcting some of the problems of standardized measurement, criterion-referenced tests are not only complex and time consuming to develop, but they often retain many of the deficiencies of standardized tests: poor attention to instructional objectives; irrelevant and misleading items; and continued use for sorting, categorizing, and predicting success rather than for diagnosing learning difficulties and improving instruction." McKenna, *A Tale of Testing in Two Cities*, in *The Myth of Measurability* 229 (P. Houts ed. 1977).

99. National Academy Report, *supra* note 3.

100. In addition to Chairperson Bailey, the committee members were John B. Carroll, Jeanne Chall, Robert Glaser, John I. Goodlad, Diane Ravitch, Lauren Resnick, Ralph W. Tyler, and Robert L. Thorndike.

101. National Academy Report, *supra* note 3, at 9 (footnote omitted).

testing programs are not insurmountable, and that there are strong indications that such programs can and do work.¹⁰² Another advocate of competency testing, Professor Ebel, contends that competency testing can help to reverse "two generations of warm-hearted but soft-headed pedagogy" that relegated "the cultivation of cognitive competence" to a secondary role. "Such tests can motivate efforts to learn and efforts to teach by providing specific goals and rewards."¹⁰³

Although many states and school districts are starting to adopt a more cautious stance, ETS and Professor Ebel appear to reflect the majority view as of this writing. Professor Decker Walker believes that support for competency testing among the official lay bodies controlling most state and local educational agencies is so strong that unanimous professional opposition, even if it could be mobilized (and it obviously cannot), would make little difference. Therefore, Professor Walker recommends that educators who see more dangers than benefits in competency testing programs should devote their efforts "to avoid the greatest dangers and maximize the positive consequences."¹⁰⁴ In sum, competency testing programs and their concomitant legal issues are likely to be with us for some time.

II. INAPPROPRIATE TEST CONTENT

Because competency testing is a relatively new phenomenon and most tests are not available for public inspection,¹⁰⁵ detailed consideration of test content is not possible at this time. There are indications, however, that competency tests in some states may incorporate inappropriate content.

Consider, for example, the following tasks set forth in the competency test guide developed for Florida's Palm Beach County schools.¹⁰⁶ The "Consumer Economics" part of the test guide includes the following comprehension task: "discuss the idea that just because a rich family can afford to feed, clothe and educate a large number of children, this does not mean that the world will be able to

102. Basic Skills Assessment, *supra* note 4, at 3.

103. Ebel, *supra* note 94, at 547.

104. Walker, *The Hard Lot of the Professional in a Reform Movement*, 35 *Educ. Leadership* 83, 84 (1977).

105. If the tests are not generally available as a policy matter under various freedom of information laws, attorneys for students who are denied diplomas on the basis of a test would probably seek a judicial order for disclosure, perhaps a protective order for limited disclosure. Because the political viability of the tests may depend on teachers and parents inspecting them to see if they are measuring things of value, some commentators recommend that tests be kept secure before but not after they are administered. Policy Issues, *supra* note 74, at 63-65. "Attempts to cut down on testing costs by having a completely secure test(s) to be used year after year therefore is probably not possible or desirable." *Id.* at 64.

106. Palm Beach County Schools, A Note Concerning APL Goals, Objectives, and Tasks (1975). The Palm Beach County schools contracted with ACT for an APL test and planned to use the test as a diploma requirement one year before the June 1979 statutory mandate.

support their children and grandchildren."¹⁰⁷ One task in the "Occupational Knowledge" part asks students "to discuss proper behavior and attitudes for keeping a job."¹⁰⁸ The "Health" part includes the following task: "discuss the physical and psychological benefits gained when food is served attractively in a pleasant atmosphere."¹⁰⁹ And one task in the "Government and Law" part asks students to discuss the concept of party politics including why the two-party system has been successful."¹¹⁰

The following "behavioral objectives" are being considered in another state¹¹¹ as part of a competency test to be used as a prerequisite to a high school diploma:

- Accept personal involvement in the Law (report crime, serve jury duty);
- State, interpret, and obey laws;
- Describe appropriate and effective ways to change laws;
- Explain the importance of basic social values (honesty, pride in work, dignity);
- Describe ways to practice responsible citizenship;
- Demonstrate self-discipline in carrying out responsibilities of citizenship;
- Justify the value of individual differences in maintaining a democratic society;
- Describe the effects of the profit system on work;
- Define personal qualities needed for work success (loyalty, responsibility, dependability);
- Demonstrate a positive self-concept;
- Understand human sexuality (physical, emotional, social factors);
- Name effective contraceptives;
- Write voluntarily to entertain self (creative language, captions, poems);
- Develop feelings of empathy and objectivity about the arts;
- Justify the value of leisure activities to others;
- Demonstrate behaviors appropriate to leisure activity.

These are but a few examples of proposed behavioral objectives that indicate test content which may be inappropriate. Detailed consideration will have to await the actual test items—that is, the specific questions and especially the answers which are deemed to be correct and incorrect. Tests will be legally vulnerable if they include items that involve (1) coerced belief, (2) invasion of privacy, (3) unteachable or unmeasurable content, (4) content that is not sufficiently matched with a school's curriculum and instruction, and (5) content that is culturally biased. While the first three kinds of inappropriate test content discussed in this Part are likely to be an immediate issue in relatively few

107. *Id.* at 28.

108. *Id.* at 31.

109. *Id.* at 10.

110. *Id.* at 34.

111. These behavioral objectives are from materials submitted to the author with an understanding of partial confidentiality. Because many states may include similar areas in subsequent stages of testing, it is more important to identify the potential problems than the particular state.

states,¹¹² the latter two are likely to be at issue in most states, and are discussed in Parts IV and V(C) respectively.

A. *Coerced Belief*

Many of the performance objectives listed earlier appear to assume that certain beliefs (behaviors, feelings, attitudes, opinions, and values generally) are the correct and proper ones, and that students can fail a test and be denied a high school diploma if their answers do not conform to those officially deemed correct. This kind of official bias would be questionable even without imposition of a serious penalty for failure to provide the "proper" answers, but combined with such a penalty the test would almost certainly constitute the kind of coerced belief prohibited by the first amendment of the Constitution.

Perhaps the most vivid expression of the first amendment's prohibition against coerced belief is found in *West Virginia Board of Education v. Barnette*,¹¹³ in which the Supreme Court held the compulsory flag salute in public schools to be unconstitutional. "The State asserts power to condition access to public education on making a prescribed sign and profession and at the same time to coerce attendance by punishing both parent and child. The latter stand on a right of self-determination in matters that touch individual opinion and personal attitude."¹¹⁴

In resolving this conflict in favor of self-determination, Justice Jackson noted the special responsibility of boards of education "for scrupulous protection of Constitutional freedoms of the individual, if we are not to strangle the free mind at its source and teach youth to discount important principles of our government as mere platitudes."¹¹⁵ The reality of a free mind is obviously incompatible with compelled belief. Thus Justice Jackson wrote: "[T]he State may 'require teaching by instruction and study of all in our history and in the structure and organization of our government, including the guarantees of civil liberty, which tend to inspire patriotism and love of country,'"¹¹⁶ but such teaching is clearly distinguishable from compulsion of students to declare a belief as is manifested by the compulsory flag salute. The flag salute is unconstitutional because it "invades

112. Most states plan to limit the scope of competency testing, at least initially, to either basic literacy and numeracy skills or to functional literacy. According to a survey by Joan Baratz, "29 states test for minimum competency in basic skills, 9 test for functional literacy and 11 have other areas in their testing programs, including citizenship, leisure skills, life long learning and attitudes toward school." *Educ. Daily*, Oct. 17, 1978, at 5.

113. 319 U.S. 624 (1943).

114. *Id.* at 630-31.

115. *Id.* at 637.

116. *Id.* at 631 (quoting *Minersville School Dist. v. Gobitis*, 310 U.S. 586, 604 (1940) (Stone, J., dissenting)).

the sphere of intellect and spirit which it is the purpose of the First Amendment to our Constitution to reserve from all official control."¹¹⁷ The Court applied "the limitations of the Constitution with no fear that freedom to be intellectually and spiritually diverse or even contrary will disintegrate the social organization."¹¹⁸

Justice Jackson further wrote:

[F]reedom to differ is not limited to things that do not matter much. That would be a mere shadow of freedom. The test of its substance is the right to differ as to things that touch the heart of the existing order.

If there is any fixed star in our constitutional constellation, it is that no official, high or petty, can prescribe what shall be orthodox in politics, nationalism, religion, or other matters of opinion or force citizens to confess by word or act their faith therein.¹¹⁹

So too any competency testing program by which the state imposes a penalty for failure of students to conform to specified beliefs would raise serious constitutional questions. Such a program would, in Justice Jackson's words, invade "the sphere of intellect and spirit which it is the purpose of the First Amendment to our Constitution to reserve from all official control."¹²⁰

Four points are worth emphasizing with respect to the *Barnette* case. First, although the plaintiffs in *Barnette* were Jehovah's Witnesses, the prohibition against coerced belief is not based upon or limited to religious objections. Justice Jackson specifically stated: "Nor does the issue as we see it turn on one's possession of particular religious views or the sincerity with which they are held."¹²¹ Thus, when he later defined the "fixed star in our constitutional constellation," that definition is broad enough to proscribe official orthodoxy in "politics, nationalism, religion, or other matters of opinion."¹²² Of course, there may be test items that meet other constitutional standards, but are constitutionally invalid solely because they violate the establishment or free exercise of religion clauses of the first amendment.¹²³

Second, test items are not rendered valid simply because the explicit or implicit values are "good" or innocuous values beyond the reach of political or other controversy. Thus, Justice Jackson stated with re-

117. *Id.* at 642.

118. *Id.* at 641.

119. *Id.* at 642.

120. *Id.*

121. *Id.* at 634.

122. *Id.* at 642.

123. For legal standards used in applying the religion and establishment clauses in school settings, see: *Epperson v. Arkansas*, 393 U.S. 97 (1968) (religion in curriculum); *School Dist. v. Schempp*, 374 U.S. 203 (1963) (Bible readings); *Engel v. Vitale*, 370 U.S. 421 (1962) (official prayers).

spect to the compulsory flag salute: "Whether the First Amendment to the Constitution will permit officials to order observance of ritual of this nature does not depend upon whether as a voluntary exercise we would think it to be good, bad or merely innocuous."¹²⁴ Under the *Barnette* rationale, the constitutional infirmity lies in the coercion, not the validity of the belief itself.

Third, the constitutional infirmity of a test cannot be cured by any consensual process such as a majority vote by the legislature, school board, parents, or students. As the *Barnette* Court noted, "[t]he very purpose of a Bill of Rights was to withdraw certain subjects from the vicissitudes of political controversy, to place them beyond the reach of majorities and officials"¹²⁵ The state legislature, school boards, community, and others who may be responsible for designing a competency testing program must do so within the limits set by the Constitution.

Finally, the principles enunciated in *Barnette* are not absolute. Although Justice Jackson stated, after his famous "fixed star" metaphor that "[i]f there are any circumstances which permit an exception, they do not now occur to us,"¹²⁶ he immediately added a footnote distinguishing situations involving military discipline.¹²⁷ The application of the *Barnette* standards will vary somewhat according to context and circumstance. An example is the landmark case of *Tinker v. Des Moines Independent Community School District*.¹²⁸

In *Tinker*, the Supreme Court considered the disruptive potential of student expression in determining whether school officials could penalize students for expressing their opposition to the Vietnam War. The *Tinker* Court referred to Justice Jackson's fixed star language with apparent approval, and held that school action prohibiting the students from wearing black armbands in school to protest the Vietnam War was a violation of the students' first amendment rights because there was no evidence that such action was necessary to avoid material and substantial interference with discipline or school work. The language of the Court underscores the constitutional prominence of free expression:

[I]n our system, undifferentiated fear or apprehension of disturbance is not enough to overcome the right to freedom of expression. Any departure from absolute regimentation may cause trouble. . . . But our Constitution says we must take this risk . . . and our history says that it is this sort of hazardous freedom—this kind of openness—that is the basis of our national strength and of the independence and vigor of Americans who grow up and live in this relatively permissive, often disputatious, society.¹²⁹

124. 319 U.S. at 634.

125. *Id.* at 638.

126. *Id.* at 642.

127. *Id.* at 642 n.19.

128. 393 U.S. 503 (1969).

129. *Id.* at 508-09 (citation omitted).

Denial of a high school diploma for failure to provide "correct" answers on a competency test obviously presents a different factual situation than exclusion from school for refusing to salute the flag or for wearing black armbands. Nevertheless, it is unlikely that the state's interest in these kinds of questions on a competency test is strong enough to justify infringement upon the individual's interest in free expression or, as discussed in the next section, privacy.

B. *Invasion of Privacy*

Some of the listed performance objectives might also be legally vulnerable because they invade personal privacy. The Supreme Court has held that a right to privacy is inherent in many of the first nine amendments to the Constitution.¹³⁰ In *Merriken v. Cressman*,¹³¹ a federal district court invalidated a test instrument on privacy grounds, stating that the right to privacy "should be treated with as much deference as free speech."¹³² The public school defendants in *Merriken* were enjoined from administering a test questionnaire designed to identify potential drug abusers. The questions asked students for personal information about themselves, their families, and other students. Although the factual situation in *Merriken* is different from that involved in competency testing, the case is instructive in suggesting some privacy limitations on test instruments that are administered to public school students.

The court found fault with the highly personal nature of the questionnaire,¹³³ the suspect results gained by the testing, and especially with two questionable uses of the test results. The court was concerned that the test results could be used by law enforcement officials who have subpoena power.¹³⁴ Moreover, it was concerned that the test results would be used for questionable remedial purposes, specifically "to change the cognitive and affective domains of potential drug abusers and other forms of deviant behavior"¹³⁵ by compulsory programs which seriously infringe upon individual liberty. Applying a balancing test involving the right of an individual to privacy on the one hand, and the right of the government to invade that privacy for the sake of the public interest on the other, the court found: "As the Program now stands the individual loses more than society can gain"¹³⁶ "There is too much of a chance that the wrong people for

130. *Roe v. Wade*, 410 U.S. 113 (1973); *Griswold v. Connecticut*, 381 U.S. 479 (1965).

131. 364 F. Supp. 913 (E.D. Pa. 1973).

132. *Id.* at 918.

133. *Id.*

134. *Id.* at 916.

135. *Id.*

136. *Id.* at 921.

the wrong reasons will be singled out and counselled in the wrong manner."¹³⁷

Some of the listed performance objectives suffer from legal infirmities similar to those identified in *Merriken*. For example, a student might refuse to acknowledge on a competency test what the authorities deem to be responsible citizenship, appropriate ways to change a law, positive self-concept, proper behavior for work or leisure situations, the success of the two-party system, or feelings of empathy and objectivity about the arts. The student might assert a privacy interest and simply refuse to disclose his or her opinions on these matters whether or not they happen to coincide with the official view. On the other hand, the student might answer the questions but assert a free expression right to have different opinions regarding the value judgments inherent in the test items.

In addition to the potential illegality of involuntary collection of information about a student's "deviant" attitude, the use of that information may also be illegal. Apart from misuse, if the school cannot protect the confidentiality of the information, there is the ominous prospect that, as in *Merriken*, classes may be created "to remedy the problem." There is also the possibility with many competency programs that students will be denied a high school diploma if the remedial attempts are unsuccessful.

C. *Unteachable or Unmeasurable Content*

Tests that go beyond basic skills and attempt to measure "the affective aspects of social responsibility, good citizenship, self-concept, and job preparedness"¹³⁸ run a high risk of violating a student's free expression and privacy interests. Such tests are also legally vulnerable not only because it is questionable that these "skills" can be measured but also questionable whether they can be generally taught given the current state of instructional and curricular research. Thus, Madaus and Airasian write:

While cognitive competencies may be taught and generally attained, teaching the affective aspects of social responsibility, good citizenship, self-concept, and job preparedness to all pupils in a school calls for knowledge and techniques simply unavailable given the current technology of instructional and curricular research . . . [O]ne continually encounters behaviors such as social responsibility, career preparedness and personal development described as "skill areas," with the implication that these behaviors can be taught in much the same way that one teaches pupils to recall number facts, balance a checkbook or read with comprehension.¹³⁹

Madaus and Airasian make a similar point about the state of the art

137. *Id.*

138. Madaus & Airasian, *supra* note 15, at 81-82.

139. *Id.*

with respect to the measurement of affective behaviors.¹⁴⁰

Punishment for something that is not one's personal fault violates a cardinal principle of liberty. Punishment, such as denial of a diploma, for failure to learn that which is not teachable or measurable raises serious legal questions. This legal concept is discussed herein with respect to tests that measure knowledge and skills not taught in the classroom.¹⁴¹

If a decision is made to include social responsibility, job preparedness, and similar areas in a competency testing program, the performance objectives and test items should be designed to elicit a student's knowledge (for example, know, describe, list, explain, apply) rather than opinions or values (for example, accept, value, appreciate, support, justify). Many of the listed performance objectives that have a value orientation could be easily revised in knowledge terms. For example, "accept personal involvement in the law" could be revised to read "explain society's (the state's) interest in personal involvement in the law." The objective is to avoid value judgments to the extent possible,¹⁴² or at least to avoid requiring implied acceptance of the value in the answer deemed to be correct. There is an obvious and crucial difference between asking a student to "describe what most employers consider appropriate in a work situation," and "describe what is appropriate behavior in a work situation."¹⁴³ The latter uses the knowledge term "describe," but is obviously written in value rather than knowledge terms.

III. ADEQUACY OF PHASE-IN PERIODS

A. Issues

Many competency programs are being imposed upon students late in their secondary education with little prior notice. Imposing the requirement one year before graduation means that students will have spent their first ten or eleven years in the school system without notice or knowledge that passing a competency test would be a condition for acquiring a diploma. In fact, the school district would have explicitly approved their progress by promoting them each year, even if they did not have basic competencies. It is likely that many if not most of those students failing the test would have studied differently in earlier years had they been given such notice—and teachers might have taught differently as well. The competency test is designed to ensure that

140. *Id.* at 87.

141. See notes 164-67 *infra* and accompanying text.

142. It is not possible, even if desirable, to design a test or run a school totally free of value judgments.

143. These examples are for general illustrative purposes. Actual test items would obviously not be written at this level of generality.

minimal competency is acquired after twelve years of schooling, but students in this situation would not have received notice until their tenth or eleventh year of schooling.

The exact date of notice will vary from school to school. In most districts there will be one date when students are given general notice of the competency requirement for a diploma, and a later date when students are given notice of the specific performance objectives to be measured by the competency test. Students, parents, and teachers should be given notice of these performance objectives and the assessment procedures as soon as possible after they have been established.¹⁴⁴

The complaint in *Debra P. v. Turlington*, challenging the Florida competency testing program, alleged in part:

Plaintiffs and members of the class they seek to represent received inadequate notice to prepare for the test. The enabling legislation for the testing program was passed in 1976. State test standards were not established until April, 1977. The first administration of the test occurred in October, 1977 and the plaintiffs herein failed.

. . . .

Defendants have changed the requirements for awarding high school diplomas after plaintiffs and members of the class they seek to represent had completed more than ten of their twelve years of public schooling.¹⁴⁵

Two students from another Florida school district make the point more bluntly. "When we got to school, they told us we have to have 20 credits to graduate," said Willie Baker. "Now they jump up and tell us we have to pass this test to graduate." Said Carrie Howard: "They should be teaching us a lot of things on the test, but they haven't given it to us yet. A person can't learn something in as short a time as they've given us."¹⁴⁶

Educators generally agree that the crucial time for teaching, learning, and remedying basic skills is in the early years of schooling. Accomplishing remedial objectives at the secondary level is more difficult and expensive than in the elementary years, and, therefore, a sound competency program should focus on the early years of schooling. The \$10 million appropriated by the Florida legislature for remedial instruction translates into only \$246 for each student who

144. Performance objectives help teachers and students focus on the knowledge and skills to be measured by the test. Some educators suggest that the test item specifications should also be designed and used for instructional and remedial purposes. Test item specifications are based upon performance objectives, but are usually more detailed in order that specific test items can be developed. Because item specifications define what the test items measure and how it is measured, they could be integrated with instructional efforts without disclosing the exact test questions. Item specifications, however, are not often subject to public disclosure and are therefore unavailable for instructional use.

145. Complaint, *supra* note 63, at ¶¶ 36, 54. See generally *id.* at ¶¶ 52-57.

146. Tallahassee Democrat, Sept. 30, 1977, § A, at 1, col. 1.

failed the October 1977 test.¹⁴⁷ As emphasized in a study of the Florida program sponsored by the Florida Teaching Profession and the National Education Association:

It appears as if the current class of eleventh graders who are Black and poor were sacrificed for the purpose of rapid implementation of the functional literacy segment of the Accountability Act.

. . . . The critical issue is whether short-term remediation programs can be effective in providing to those poor and Black children knowledge and skills which the schools have not been successful in imparting over the last 11 years.¹⁴⁸

The point, of course, is not to withhold remedial efforts at the secondary level, but rather to question the fairness and legality of withholding diplomas where belated efforts are not successful.

One of the most important decisions for policymakers who are committed to this kind of program is setting the initial date when diplomas will depend upon the results of a competency test. Problems of adequate notice and adequate match between test and instruction will probably correct themselves in the long term. The task for policymakers is to make sure that whole generations of low-scoring students are not sacrificed in the short term for the sake of rapid program implementation.

B. Analysis

Traditional notions of due process should require adequate prior notice of any rule that could cause irreparable harm to a person's educational or occupational prospects.¹⁴⁹ *Mahavongsanan v. Hall*,¹⁵⁰ a

147. These figures are based upon the 40,700 students who failed the October 1977 test. As of this writing, there is a bill in the Florida legislature to increase the amount appropriated for remedial instruction to \$26 million (approximately \$640 per student based on 40,700 students).

148. National Education Association, *Impact of Minimum Competency Testing in Florida*, Today's Educ., Sept.-Oct., 1978, at 35-36. The cited article reprints excerpts from a report prepared by a study panel consisting of Ralph W. Tyler (Chairperson), Stephen Lapan, Judy Moore, L. Wendell Rivers, and Donna Skibo. The study panel, sponsored by the Florida Teaching Profession and the National Education Association (NEA), held hearings in Florida in the spring of 1978. The panel identified "several serious defects in implementation" that appeared to be "largely the effects of rushing to establish the program without considering all the important aspects of it." *Id.* at 31.

149. The legal standard applied in modern substantive due process cases (under the fourteenth amendment of the Constitution or comparable state due process provisions) is usually not carefully set out, but state action is usually illegal if it (1) is arbitrary or capricious, (2) does not achieve any legitimate state interests, (3) frustrates any legitimate interest the state might have, or (4) is fundamentally unfair. See McClung, *The Problem of the Due Process Exclusion: Do Schools Having a Continuing Responsibility To Educate Children with Behavior Problems?*, 3 J. L. & Educ. 491, 495-501 (1974). Whatever the exact wording of the test under substantive due process, the general standard is that state action cannot be unreasonable, with unreasonableness being construed narrowly (e.g., rational persons would not disagree). A stronger standard has been applied in some situations. See *St. Ann v. Palisi*, 495 F.2d 423 (5th Cir. 1974).

150. 529 F.2d 448 (5th Cir. 1976).

Fifth Circuit decision, provides some support for the proposition that students must be provided with adequate notice of any significant change in graduation requirements, even though the court found that the plaintiff in that case had not been denied procedural or substantive due process because she received "timely notice" that passing a comprehensive examination would be a prerequisite to a master's degree in education.¹⁵¹

The legal argument for adequate notice of significant changes in graduation requirements is stronger in the context of elementary and secondary public schools than in postsecondary education because courts usually apply a stricter standard of review to practices at the elementary and secondary level where education is compulsory. Notice of a competency test requirement for a high school diploma would have to be much earlier than in *Mahavongsanan* because twelve rather than just two or three years of education are being tested. The plaintiff in *Mahavongsanan* received notice of the new requirement only six months after starting the program. Whatever notice is considered adequate for the competency testing requirement, notice after most of one's educational program is already completed seems clearly inadequate, especially if the competency test is designed to measure knowledge and skills not previously taught in the district's classrooms.

IV. MATCHING TEST AND INSTRUCTION

Most persons would agree that fairness requires a school's curriculum and instruction to be matched in some way with whatever is later measured by the competency test. In other words, the test would be unfair if it measured what the school never taught. This concept should be considered in terms of curricular and instructional validity.

A. *Curricular and Instructional Validity*

Curricular validity is a measure of how well test items represent the objectives of the curriculum.¹⁵² An analysis of curricular validity would require comparison of the test objectives with the school's course objectives. For example, if the curriculum was not designed to teach functional competency, it would be unfair to deny students their diploma because they did not learn to be functionally competent. In this situation, failure on the competency test should reflect on the schools, which are not offering an appropriate curriculum.

A competency test should also have what may be called *instructional validity*.¹⁵³ Even if the curricular objectives of the school correspond

151. *Id.* at 450.

152. See Cronbach, *Essentials of Psychological Testing* 397 (1970); Dick & Hagerty, *Topics in Measurement: Reliability and Validity* 96 (1971).

153. The discussion in this Article of the various types of test validity was developed with the assistance of Diana Pullin of the Center for Law and Education. Because neither the APA

with the competency test objectives, there must be some measure of whether the school district's stated objectives were translated into topics actually taught in the district's classrooms. While a measure of curricular validity is a measure of the theoretical validity of the competency test as a test instrument to assess the success of students, instructional validity is an actual measure of whether the schools are providing students with instruction in the knowledge and skills measured by the test. Instructional validity obviously does not require prior exposure of the student to the exact questions asked on the test,¹⁵⁴ but it does require actual exposure of students to the kind of knowledge and skills that would enable a student to answer the test questions. This will present difficult problems of proof in some cases; in others, it will be relatively easy to show that the test is measuring what the school never taught. Some possible approaches to assessing the instructional validity of a test are discussed in Part VII(C).

It is important to note that content validity, as defined by the APA,¹⁵⁵ does not ensure either curricular or instructional validity. The concepts are related but distinguishable. Content validity is a measure of how well test items represent the performance domain that the test purports to measure (for example, adult performance skills), but it is not necessarily a measure of how well the test items and performance domain represent either a particular school's curricular objectives or instruction received. Instructional validity should be the central concern, because content and curricular validity mean very little in this context if the test items are not representative of instruction actually received by the student.

Test developers will be reluctant to evaluate the curricular or instructional validity of their tests, but they have some kind of professional responsibility (beyond the usual disclaimers) to ensure that their tests do not measure skills and knowledge that were never taught in school, particularly when they design a test instrument for an individual user. The ultimate responsibility for the curricular and

Standards nor the educational literature generally seems to address the issue of matching the test with actual instruction (especially important in the fair administration and assessment of any minimal competency program), we developed the concept of "instructional validity" for this purpose.

154. This raises the classic controversy about "teaching to the test." See Madaus & Airasian, *supra* note 15, at 82-90; note 144 *supra*.

155. APA Standards, *supra* note 67, at 28. The general nature of the discussion of content validity in the APA Standards is perhaps explained by the heavy focus on employment testing. The APA Standards were developed during a period when competency testing as a prerequisite to a high school diploma was not an issue. Curricular and instructional validity are irrelevant concepts in most employment tests. Teachers certification tests, however, are an important exception. See, e.g., *United States v. South Carolina*, 15 Fair Empl. Prac. Cas. 1196 (D.S.C. 1977), *aff'd mem. sub. nom. National Educ. Ass'n v. South Carolina*, 434 U.S. 1026 (1978). In fact, the APA Standards were not written with the intent that they have the force of law. Novick, *supra* note 68, at 47. Ideally, the proposed revision of the APA Standards will incorporate concepts like curricular and instructional validity for competency tests used as a prerequisite to a high school diploma or grade-to-grade promotion.

instructional validity of the test, however, should be with the school or school system that uses the test. A school or school system that cannot ensure the curricular and instructional validity of its competency tests should not use them as a basis for denying promotion or a diploma to any of its students.¹⁵⁶

Questions of curricular and instructional validity are relevant, given either of two general purposes of competency testing: (1) to measure the student's mastery of the school's curriculum or (2) to predict the minimal competence required in the adult world.¹⁵⁷ The terminology will vary from school to school, and some schools will merge the two concepts by deciding that their curricula should be based upon minimal adult competence. The different versions of adult competency tests usually go beyond basic proficiency skills, because they seek to measure an individual's ability to apply basic skills (including literacy) to necessary adult-life role activities such as those of a consumer, producer, or citizen. Examples include ability to understand common indices for comparison shopping, to understand the nutritional ingredients necessary for a balanced diet, to understand a contract for a car loan or home mortgage, to read and understand a newspaper, to fill out a job or loan application, to complete a tax form, to balance a checkbook, to follow a recipe in preparing a meal, to understand proper behavior and attitudes for getting and keeping a job, to use leisure time productively, and to participate as a citizen in the community, state, and nation.¹⁵⁸

B. *Basic Literacy Skills Versus Adult-Life Skills*

Some people do not think that the distinction between basic literacy/numeracy skills and adult-life skills is an important one since they believe that a student with basic literacy skills will have no difficulty passing an adult-life skills test even if he or she has not been exposed to adult-life skills in the classroom. They argue that students who can add and subtract a series of four-digit numbers on a basic numeracy skills test item, for example, will be able to do the same in the context of a tax form item on an adult-life skills test. However, many students, especially low-scoring students, will have difficulty with the kind of transference skills called for in an adult-life skills item.¹⁵⁹ In fact, many of these students probably score low because

156. For a discussion of curricular and instructional validity in the context of competency testing, see: Policy Issues, *supra* note 74, at 67-71; pt. VII(C) *infra*.

157. If the purpose of the proficiency test is to measure the minimal competence required in the adult world, the test should also have predictive validity or an acceptable substitute. See Policy Issues, *supra* note 74, at 73, 74; note 74 *supra* and accompanying text.

158. Some of these examples suggest inappropriate test content. See pt. II *supra*.

159. "One system in Massachusetts completed a study documenting that students had learned traditional school skills, but needed work in the area of transferring or applying of those skills to everyday life." Minimum Competency Testing, *supra* note 2, at 26; see Office of Program

they have difficulty with transference. Therefore, school districts that plan to test for adult-life skills should have curricula and instruction that emphasize transference as well as the other knowledge and skills necessary to answer the adult-life skills items.

An adult-life skills item, such as adding and subtracting four-digit numbers on a tax form, is also more difficult than its basic numeracy components for other reasons, including the fact that (1) the forms usually require literacy as well as numeracy skills, and (2) an official form can be distracting and intimidating, especially if the students are not familiar with the form. Of course, it is possible to write a basic literacy or numeracy skill item which is more difficult than its adult-life skills counterpart or to make an adult-life skills test easier by allowing a lower cut-off score than for the literacy/numeracy skills test. But in general, an adult-life skills test designed to measure the same literacy and numeracy skills as a basic skills test will be more difficult for most students; and a school's curricula and instruction should account for this difference.

Schools may not have taught the life skills that are measured by many competency tests. "Interest rates, checkbooks, tax forms, etc., were included in the 7-8 grade arithmetic books in the 40's and 50's but dropped out of sight in the 60's with the advent of modern math, the exception being courses like business arithmetic where such topics are still covered."¹⁶⁰ There may be a mismatch between test and instruction even where the test is limited to basic literacy and numeracy skills. For example, some schools may not have taught or emphasized some basic skills such as metric conversion that are measured by many competency tests. Even where basic skills are part of the curriculum, Popham notes that instructional differences surface when such general concepts as "the ability to perform fundamental mathematical operations"¹⁶¹ are translated into specific test items.¹⁶² This problem will be especially acute where (1) the state uses a single statewide test that does not account for the diverse curricula and instruction developed in individual school districts, or (2) an individual district purchases a standardized test that was developed without regard to local curriculum and instruction.¹⁶³

Where an adult competency purpose is involved, as is the case in

Evaluation and Research, California State Dep't of Education, *Getting Inside a Student Competency Test 7* (1978).

160. Policy Issues, *supra* note 74, at 69. For a more complete description of the ways in which conventional curriculum and instruction are not well matched with competency based tests, see Spady & Mitchell, *Competency Based Education: Organizational Issues and Implications*, Educ. Researcher, Feb., 1977, at 9-10, 13.

161. Popham, *supra* note 81, at 471.

162. *Id.* at 470-71.

163. See note 261 *infra* and accompanying text.

many school systems where various adult performance level tests have been adopted, it is less likely that the school will in fact have taught what is measured by the test. If this is the case, greater curriculum revision and longer phase-in periods will be necessary. A competency test that measures adult-life role skills that were never taught in the school and then is used as a basis for denying a diploma is arguably so arbitrary as to violate due process of law.¹⁶⁴ A competency test lacking curricular or instructional validity might violate substantive due process because then the students are being penalized even though they cannot be personally faulted for poor performance on the test. Support for this argument can be found in *St. Ann v. Palisi*,¹⁶⁵ in which the Fifth Circuit held that a school board regulation violated substantive due process because it allowed school children to be suspended for their parent's misconduct. In so holding, the court said: "[T]he children do not complain that they were denied the constitutional right to an education, but that they were punished without being personally guilty. Thus a cardinal notion of liberty is involved and substantive due process is applicable."¹⁶⁶ Since the practice established "a significant encroachment upon a basic element of due process," the court required the school to meet "a substantial burden" to justify this encroachment, including proof that reasonable alternative means to achieve the stated objective were not available.¹⁶⁷

In another analogous case, a three-judge court in *United States v. South Carolina*¹⁶⁸ held, *inter alia*, that a study conducted by ETS for South Carolina, to demonstrate that the content of the National Teacher Examination (NTE) matched the content of teacher training programs in the state, was sufficiently trustworthy to sustain the state's burden under Title VII of the Civil Rights Act of 1964¹⁶⁹ to show the validity of the test, even though the validation was not related to a particular school's instruction¹⁷⁰ or to job performance. With Justices White and Brennan dissenting, and Justices Marshall and Blackmun not taking part, the Supreme Court affirmed the decision of the lower court without writing an opinion.¹⁷¹ Since this case involved employment testing for purposes of hiring and classifying teachers, it does not provide a firm basis for prediction about judicial review of validation studies of competency tests used for purposes of awarding high school diplomas.

164. See note 149 *supra* and accompanying text.

165. 495 F.2d 423 (5th Cir. 1974).

166. *Id.* at 426-27.

167. *Id.* at 427.

168. 15 Fair Empl. Prac. Cas. 1196 (D.S.C. 1977), *aff'd mem. sub nom.* National Educ. Ass'n v. South Carolina, 434 U.S. 1026 (1978).

169. 42 U.S.C. §§ 2000a to 2000h-6 (1976).

170. The adequacy of the match between the test and a particular school's instruction (that is, instructional validity) was not at issue in this case.

171. National Educ. Ass'n v. South Carolina, 434 U.S. 1026 (1978).

Plaintiffs in *Debra P.* include in their complaint the allegation that the Florida statutes have never required school districts to provide, and the defendant school district had not in fact provided, instruction and curriculum in all areas covered by the functional literacy test.¹⁷² They allege that defendants' failure to ensure that there is a match between the instruction and curriculum provided in the classroom and the areas covered by the test constitutes the kind of arbitrary, capricious, and fundamentally unfair action prohibited by the due process guarantee of the fourteenth amendment.¹⁷³

In sum, a number of factors, such as statewide tests and tests of adult-life skills, suggest that many competency tests may measure what many students have not been taught. The need for some kind of match between test and instruction seems basic to the concept of fair play, and an initial question will be whether the test in fact reflects curriculum and instruction. Whether courts will require instructional and/or curricular validity of competency tests that are used as a prerequisite to a high school diploma is a separate, unanswered question. If so, the requisite degree of match between test and instruction and/or curriculum,¹⁷⁴ and the kind of evidence used to determine that match,¹⁷⁵ are additional, unanswered questions.

V. RACIAL AND LINGUISTIC DISCRIMINATION

While substantial numbers of white middle-class students cannot meet minimal competency standards, some evidence indicates that a disproportionate number of black and Hispanic students will not be able to meet the competency test requirements. This pattern has in fact occurred where competency testing programs have been implemented.¹⁷⁶ For example, the statewide assessment results released by the Florida State Department of Education on March 16, 1978 showed that approximately 75% of the state's black students (juniors) had failed the mathematics part of the statewide functional literacy test. Approximately 40% of Hispanic, 30% of Asian, and 25% of white students also failed the mathematics part of the test.¹⁷⁷ Results of

172. Complaint, *supra* note 63, at ¶¶ 37-38.

173. *Id.* at ¶¶ 56-57; see note 149 *supra*.

174. See notes 258-62 *infra* and accompanying text.

175. Evidence of instructional validity might include teacher testimony, homework samples, and materials used in the classroom. See notes 262-63 *infra* and accompanying text. Some commentators are concerned that instructional validity might necessitate more extensive record-keeping procedures. See Policy Issues, *supra* note 74, at 69-71; Duke, Donmoyer, & Farman, *Emerging Legal Issues Related to Classroom Management*, 60 Phi Delta Kappan 305, 307 (1978). See generally note 255 *infra*.

176. See note 44 *supra*.

177. Florida Dep't of Education, Statewide Assessment Results by Demographic Categories,

earlier testing in Florida also indicated a disproportionate failure rate among minority students.

Detailed analysis of racial and linguistic discrimination must be tailored to the specific factual situation in each school district. Basically, any school district will fall into one of the following three categories: (1) a district with a court finding of prior discrimination;¹⁷⁸ (2) a district with no such finding but vulnerable to such a finding if the issue were to be litigated; and (3) a district with no such finding since the district has provided equal educational opportunities to all of its students. There are many variations on these three situations, and each district should seek legal consultation to determine whether past practices in the district combined with its version of competency testing constitute racial or linguistic discrimination. Districts that have already been found to have discriminated, or that are vulnerable to such a finding, must be especially sensitive to the discriminatory potential of competency testing and should design their programs to account for possible problems.

A. *Racial Discrimination*

Some black parents in desegregated communities see a racial motive behind competency testing. They say that competency testing was not a concern at either black or white schools until the schools in their district were desegregated, at which time competency testing was introduced "to protect standards." The effect can be "resegregation" within the school according to test results (or other forms of tracking) since unequal educational opportunities may cause black children to score lower than their white counterparts.

Whether or not a racial motive is involved, such practices are arguably unconstitutional in formerly segregated districts. In comparable situations, the federal courts have held that practices which carry forward the effects of prior racial discrimination are prohibited. For example, in *Gaston County v. United States*,¹⁷⁹ an action brought under the Voting Rights Act of 1965,¹⁸⁰ the Supreme Court held that it was appropriate for a court to consider whether a literacy or educational requirement had the effect of denying the right to vote on account of race or color because the state had maintained separate and inferior schools for its black residents who were then of voting age. The Court stated: "[W]e cannot escape the sad truth that through-

Fall, 1977 Tests, Grade Eleven Highlights (Functional Literary) (1978). The extent to which Florida will be able to reduce this disproportionate failure rate, and the overall failure rate, by the time these students graduate is unclear as of this writing.

178. For example, the district may be under a court-ordered desegregation plan pursuant to a judicial finding of purposeful segregation and/or provision of unequal educational opportunities.

179. 395 U.S. 285 (1969).

180. 42 U.S.C. §§ 1971, 1973-1973p (1976).

out the years Gaston County systematically deprived its black citizens of the educational opportunities it granted to its white citizens. 'Impartial' administration of the literacy test today would serve only to perpetuate these inequities in a different form."¹⁸¹

Similarly, the Fifth Circuit has determined that the testing necessary for ability grouping cannot be applied to black students for the first time in the years immediately following desegregation.¹⁸² This "prior effects" principle arguably applies in other cases of serious injury, such as denial of grade promotion or a high school diploma. It forms the basis for one of the claims in the challenge to the Florida competency testing program.

In *Debra P.*, plaintiffs cite a lengthy history of deliberate racial segregation of schools in Hillsborough County and in the state of Florida generally.¹⁸³ They state that after the Florida Pupil Assignment Law¹⁸⁴ was declared unconstitutional in 1962, the county and state continued to resist the dismantling of the dual school system, as evidenced by the various federal court findings in *Mannings v. Board of Public Instruction*.¹⁸⁵ Plaintiffs further claim that the Hillsborough public schools were unconstitutionally segregated by race when the plaintiffs entered school in the first grade, and as a result they received an inferior education which contributed to their failing the test.¹⁸⁶ As a result, the plaintiffs contend that they have been re-segregated in predominantly black remedial classes, and are in imminent danger of diploma denial, permanent stigmatization, diminished employment opportunities, loss of income, and restricted access to higher education.¹⁸⁷ In sum, the plaintiffs argue that Florida's competency testing program perpetuates the effects of past racial discrimination in violation of the equal protection clause of the Constitution and Title VI of the Civil Rights Act of 1964.¹⁸⁸ In summary, the legal prohibition against carrying forward the effects of past discrimination is a prohibition against placing one injury, diploma denial, on top of another, unequal educational opportunities—in other words, "blaming the victims" for conditions caused by earlier governmental discrimination.

181. 395 U.S. at 296-97; see *Oregon v. Mitchell*, 400 U.S. 112 (1970); *Kirksey v. Board of Supervisors*, 554 F.2d 139 (5th Cir.), *rev'g en banc* 528 F.2d 536 (5th Cir.), *cert. denied*, 434 U.S. 968 (1977).

182. *United States v. Gadsden County School Dist.*, 572 F.2d 1049 (5th Cir. 1978); *McNeal v. Tate County School Dist.*, 508 F.2d 1017 (5th Cir. 1975); *Moses v. Washington Parish School Bd.*, 456 F.2d 1285 (5th Cir.), *cert. denied*, 409 U.S. 1013 (1972); *Lemon v. Bossier Parish School Bd.*, 444 F.2d 1400 (5th Cir. 1971).

183. Complaint, *supra* note 63, at ¶¶ 20-29.

184. Act of July 26, 1956, ch. 31380, 1955-1956 Fla. Laws 30 (repealed 1965).

185. 427 F.2d 874 (5th Cir. 1970).

186. Complaint, *supra* note 65, at ¶ 25.

187. *Id.* at ¶ 39.

188. *Id.* at ¶ 51.

Plaintiffs may have to prove a discriminatory purpose when challenging competency testing programs in school districts not recently desegregated or found to be subject to prior discrimination. The Supreme Court held in *Washington v. Davis*¹⁸⁹ that the disproportionate racial impact of a police department's personnel test was not sufficient to establish an unconstitutional racial classification without proof that it reflected a racially discriminatory purpose.¹⁹⁰ The Court, however, stated that such disproportionate racial impact could be evidence of a discriminatory purpose.¹⁹¹

In a concurring opinion, Justice Stevens noted that "the line between discriminatory purpose and discriminatory impact is not nearly as bright, and perhaps not quite as critical,"¹⁹² as the majority's opinion suggested. "Frequently the most probative evidence of intent will be objective evidence of what actually happened rather than evidence describing the subjective state of mind of the actor. For normally the actor is presumed to have intended the natural consequences of his deeds."¹⁹³ Given the studies cited above, the natural consequence of most competency testing programs will be racial differentiation.

The standards used to prove intent to segregate in school desegregation cases are instructive for purposes of proving discriminatory intent in competency testing cases. In *Village of Arlington Heights v. Metropolitan Housing Development Corp.*,¹⁹⁴ the Supreme Court suggested that evidence as to the following factors might have probative value in proving intent: historical background, the sequence of events leading up to the challenged decision (including substantive and procedural departures from the norm), legislative history, and testimony from officials.¹⁹⁵ The "natural and foreseeable" standard for establishing intent to segregate has been approved by at least five federal circuit courts.¹⁹⁶ The Second Circuit in *Arthur v. Nyquist*¹⁹⁷ said:

When . . . [school board] actions have the "natural, probable, and foreseeable result of increasing or perpetuating segregation," a presumption of segregative purpose is created. The burden of proof then shifts to defendant officials . . . to demonstrate that no reasonable alternative policy would have achieved the same permissible educational goals with less segregative effect.¹⁹⁸

189. 426 U.S. 229 (1976).

190. *Id.* at 246.

191. *Id.* at 242.

192. *Id.* at 254 (Stevens, J., concurring).

193. *Id.* at 253 (Stevens, J., concurring).

194. 429 U.S. 252 (1977).

195. *Id.* at 267-68.

196. See *United States v. Texas Educ. Agency*, 564 F.2d 162, 168 (5th Cir. 1977) (standard adopted by First, Second, Fifth, Sixth, and Seventh Circuits).

197. 573 F.2d 134 (2d Cir.), *cert. denied*, 99 S. Ct. 179 (1978).

198. *Id.* at 142-43.

The Fifth Circuit, in *United States v. Texas Education Agency*,¹⁹⁹ held that the "natural and foreseeable" standard for establishing segregative intent is consistent with the Supreme Court decisions in *Washington v. Davis* and *Village of Arlington Heights* which stress that otherwise neutral state action does not violate the fourteenth amendment's equal protection clause solely because it has a disproportionate impact on a racial minority. Thus the Fifth Circuit concluded: "When the official actions challenged as discriminatory include acts and decisions that do not have a firm basis in well accepted and historically sound non-discriminatory social policy, discriminatory intent may be inferred from the fact that those acts had foreseeable discriminatory consequences."²⁰⁰

In *Washington v. Davis*, the Supreme Court also distinguished between the constitutional standard and the Title VII standard on testing, noting that the latter was more stringent since it incorporated an effect rather than a purpose standard.²⁰¹ Thus when a test or practice disqualifies substantially disproportionate numbers of blacks in hiring and promotion decisions, the burden under Title VII shifts to the employer to validate the test or practice in terms of job performance and to show that the test or practice is sufficiently job-related.

The HEW regulations implementing Title VI of the Civil Rights Act of 1964²⁰² incorporate a similar effect (rather than purpose) standard, prohibiting practices which have the effect of discriminating against individuals on the ground of race, color, or national origin.²⁰³ This Title VI effect standard has been cited with approval and applied by the Supreme Court in *Lau v. Nichols*,²⁰⁴ although recently questioned.²⁰⁵ The Title VI regulations also incorporate the *Gaston County* principle against carrying forward the effects of past discrimination,²⁰⁶ and add to it an affirmative obligation to take steps to remedy those effects.²⁰⁷

199. 564 F.2d 162 (5th Cir. 1977).

200. *Id.* at 168.

201. 426 U.S. at 246-48.

202. 42 U.S.C. § 2000d (1976).

203. Title VI states: "No person in the United States shall, on the ground of race, color, or national origin, be excluded from participation in, be denied the benefits of, or be subjected to discrimination under any program or activity receiving Federal financial assistance." *Id.* Regulations issued by HEW pursuant to this section state that a recipient of federal funds "may not . . . utilize criteria or methods of administration which have the effect of subjecting individuals to discrimination because of their race, color, or national origin, or have the effect of defeating or substantially impairing accomplishment of the objectives of the program as respect individuals of a particular race, color, or national origin." 45 C.F.R. § 80.3(b)(2) (1978) (emphasis added); see *id.* § 80.3(b)(1)(iv)-(v), .5(e) (1978).

204. 414 U.S. 563, 568 (1974).

205. See notes 209-10 *infra* and accompanying text.

206. See notes 179-81 *supra* and accompanying text.

207. 45 C.F.R. § 80.3(b)(6)(i) (1978) provides: "In administering a program regarding which

Although HEW has not taken any position on competency testing as of this writing, application of these Title VI standards to public school testing programs is indicated by an HEW memorandum requiring schools to take steps "to adopt and implement procedures to insure that test materials and other assessment devices used to identify, classify and place exceptional children are *selected and administered* in a manner which is *non-discriminatory in its impact* on children of any race, color, [or] national origin"208 Because virtually all public schools are subject to Title VI regulations, competency testing programs that have a disproportionate effect on blacks or other protected minorities should be examined in light of Title VI standards, especially where there is evidence of racial bias in the test itself or in the administration of the test.

Applying Title VI standards to competency testing, however, is complicated by the Supreme Court's decision in *University of California Regents v. Bakke*.²⁰⁹ In that landmark case setting the legal parameters of affirmative action programs, Justice Powell and four other justices agreed that Title VI proscribes only those racial classifications that would violate the equal protection clause if employed by a state or its agencies. Given the Court's earlier decision in *Washington v. Davis* that a disproportionate racial impact is not sufficient to prove a constitutional violation, *Bakke* would appear to undercut the effect standard incorporated in the Title VI regulations. Justice Powell, writing for a divided Court, nevertheless indicated approval of *Lau v. Nichols*²¹⁰—a case in which no discriminatory purpose was proven, and the Title VI effect standard was used to require English-language instruction to students of Chinese ancestry who do not speak English. Therefore, the status of the effect standard incorporated in HEW's Title VI regulations will remain somewhat uncertain until further clarified by the courts. Plaintiffs challenging competency testing programs, however, should probably proceed under the assumption that they will have to prove discriminatory intent, except where the program carries forward the effects of proven, past discrimination.

B. *Linguistic Discrimination*

Similar standards apply to competency testing programs that have a disproportionate effect on Hispanic children. Some school districts are

the recipient has previously discriminated against persons on the ground of race, color, or national origin, the recipient must take affirmative action to overcome the effects of prior discrimination." For illustrations of the principle, see 45 C.F.R. §§ 80.5(i)-(j) (1978).

208. U.S. Dep't of Health, Education and Welfare, *Identification of Discrimination in the Assignment of Children to Special Education Programs 3* (1975) (memorandum by M. Gerry, Acting Director, Office for Civil Rights) (emphasis added).

209. 438 U.S. 265 (1978).

210. 414 U.S. 563 (1974).

imposing the competency requirements without a Spanish translation of the test or corresponding curricular and instructional modification. Such practices may not satisfy HEW's standards requiring public schools to take affirmative steps to remedy the linguistic exclusion of limited and non-English speaking children.²¹¹

The constitutional prohibition against carrying forward the effects of past discrimination may also apply. Consider, for example, a school district that—pursuant to *Lau v. Nichols* and related HEW memoranda²¹²—has implemented meaningful bilingual programs for limited and non-English-speaking students at the elementary grades, but has not provided adequate bilingual programs for the first group of limited and non-English-speaking students scheduled to graduate subject to the new competency test requirement. Hispanic students who cannot pass the test might argue that denial of the diploma is illegal because it simply reflects unequal educational opportunities provided by the school. Like many black students, Hispanic students in such a situation would probably argue that denying them a diploma is in effect “blaming the victims.”²¹³

In addition to the protection provided by Title VI and the Constitution, there are separate legal standards that specifically prohibit linguistic discrimination in educational programs. For example, the Equal Educational Opportunity Act of 1974²¹⁴ provides: “No State shall deny equal educational opportunity to an individual on account of his or her race, color, sex, or national origin, by . . . the failure by an educational agency to take appropriate action to overcome language barriers that impede equal participation by its students in its instructional programs.”²¹⁵

The point here is not that minorities should have to meet lesser standards of functional competence than whites. In fact, blacks and Hispanics, who in disproportionate numbers have been enrolled in schools that do not emphasize basic skills, are in the forefront of those calling for increased attention to basic skills. But care must be taken to ensure that competency testing does not lead simply to another form of racial and linguistic discrimination.

211. See U.S. Dep't of Health, Education and Welfare, Identification of Discrimination and Denial of Services on the Basis of National Origin (1970) (memorandum by J. Pottinger, Director, Office for Civil Rights).

212. See *id.*

213. To meet this concern in part, the Massachusetts Department of Education's Proposed Regulations for Basic Skills Improvement requires exemption of all students “identified, screened and enrolled in transitional bilingual programs pursuant to General Laws, Chapter 71A.” Proposed Regulations, *supra* note 29, § 40.10(4).

214. 20 U.S.C. §§ 1701-1758 (1976).

215. *Id.* § 1703(f) (1976). This statute also specifically provides for a civil remedy in federal district court. *Id.* § 1706.

C. Cultural Bias

The disproportionate racial and linguistic effect of most competency tests may result in part from cultural bias in test content. Cultural bias is a concept subject to many definitions and interpretations. Some would argue that a test is culturally biased if the results are used unfairly, while others would distinguish between cultural bias in test content and unfair use of the test. Some would define bias to encompass any situation where there is a systematic difference in test scores among groups, whereas others would limit the term to situations where there are items within a test that are especially more difficult for one group than another relative to other items. These are only a few of many possible definitions.²¹⁶ When the complex emotional and political aspects of any definition are combined with the technical measurement problems, it is unsurprising that no single definition of cultural bias has won widespread acceptance.

Defining cultural bias in the context of competency and other forms of achievement tests poses special problems. What is determined to be cultural bias in an Intelligence Quotient (IQ) or aptitude test may not be cultural bias in a competency or achievement test because the tests purport to measure different things. IQ tests purport to measure innate intelligence, and aptitude tests purport to predict an individual's potential for future achievement—and thus these tests theoretically should not be influenced by instruction.²¹⁷ Competency and other achievement tests, however, are designed to measure what skills and knowledge an individual has learned in a particular setting, and therefore should be sensitive to and reflective of instruction. Thus the specific findings regarding cultural bias in legal challenges to IQ testing²¹⁸ are not likely to resolve similar questions about bias in competency testing, although the general legal analysis should be instructive.²¹⁹

216. For a helpful summary and overview of the various definitions and models to determine bias, see Linn, *In Search of Fair Selection Procedures*, 13 J. Educ. Measurement 53 (1976). Of special note is Linn's discussion of criterion problems. See *id.* at 56-57. See generally APA Standards, *supra* note 67, at 43-46.

217. Intelligence and competency testing purport to measure different things, but many critics argue that they measure essentially the same thing—achievement, or some indistinguishable combination of inherited intellectual capacity and achievement. See, e.g., J. Schwartz, *The Illogic of IQ Tests*, in *The Myth of Measurability* 90 (P. Houts ed. 1977).

218. See *Larry P. v. Riles*, 343 F. Supp. 1306 (N.D. Cal. 1972), *aff'd per curiam*, 502 F.2d 963 (9th Cir. 1974). Plaintiffs in *Larry P.* were black children who argued that the stigmatizing EMR (educable mentally retarded) label was applied to them on the basis of culturally biased IQ tests, which penalized unfamiliarity with white middle-class background. The court concluded that defendants had not sustained their burden of proving a rational relationship between scoring on the questioned IQ test and the ability of black students to learn, and, therefore, awarded preliminary injunctive relief against future testing. *Id.* at 1314. Although the trial on the merits was concluded in the spring of 1978, no decision has been handed down as of this writing.

219. Of particular interest, for example, is the legal standard that will be applied by the

A "culture-fair" test may be an impossible goal. Henry C. Dyer of ETS maintains that there are only two conditions under which a test can be culture-fair: (1) either the learning required to perform acceptably on the test is commonly and equally available to all people of all cultures, or (2) the stimulus material on the test is novel to all people of all cultures. He contends that neither one of these conditions is obtainable.²²⁰ This sobering realization, however, should not hinder the development of tests which are as bias-free and culture-fair as possible for those situations where testing is required as part of an educational program.

Since most competency tests purport to measure school achievement and to be sensitive to instruction, such tests should have instructional validity.²²¹ Serious questions of test bias would be raised, for example, if the competency test measured knowledge and skills taught to white but not to minority students. The more difficult questions, however, are those raised by test items that are more difficult for minority students, in spite of being offered the same instruction, because of their different cultural backgrounds. The point is dramatically made by student performance on tests which are designed to measure intelligence and knowledge familiar to students from minority cultures.²²² Not surprisingly, minority students outperform most white middle-class students who are at a disadvantage on these tests because of their background.

One possible solution to this problem is to try to design a test that reflects all aspects of our pluralistic society—or at least the extent of diversity reflected by the student population. This would have the effect of equalizing the test by placing students at an equal advantage or disadvantage. This approach also has the advantage of being a more accurate measure of the skills needed to function effectively in a pluralistic society. A functional competency test given in Miami or San Antonio, for example, should include a number of Hispanic skill and content items, as some cross-cultural competence is arguably necessary for successful functioning in those cities.

The technical difficulties involved in designing a fair pluralistic test lead some persons to favor and the goal of "culturally neutral" tests. The problem, of course, is that a culturally neutral test is a more evasive goal than a pluralistic test. Tests usually perceived as culturally

Larry P. court in light of the Supreme Court's decision in *Univ. of Cal. Regents v. Bakke*, 438 U.S. 265 (1978).

220. H. Dyer, *Race and Intelligence: An Examination of the Scientific Evidence by Four Authorities* (1963).

221. See pt. IV(A) *supra*.

222. An example of this kind of test is *The Black Intelligence Test of Cultural Homogeneity (BITCH)-100—A Cultural Specific Test*, designed by R. L. Williams and available from Williams & Associates, Inc., 6374 Delmar Blvd., St. Louis, Mo. 63130.

neutral in our society are almost inevitably tests that favor the predominantly white middle-class culture, in other words, the majority culture.²²³ Another argument against a pluralistic test is the practical one that success is defined by the majority culture, and all students regardless of background should be provided with the education and training necessary to function effectively in our predominantly white middle-class culture. Standard English and other values of the majority culture are the ultimate indicators of success, according to this argument, and schools should help all students to achieve this goal. This practical argument is voiced by many minority parents and students.

If majority culture values are assumed, then the question of fairness remains, given the inherent advantage that majority culture students have over minority students even on instructionally valid tests. One approach would be to minimize minority disadvantage on these tests by providing additional, special instruction in majority-culture aspects of the test that are especially difficult for minority students. Bilingual instruction for limited and non-English speaking students is the most obvious example, and comparable programs for other minority students could be developed.

Pretesting can help to determine which aspects of a competency test are especially difficult for minority students. Several statistical approaches have been developed to determine which items in a test are especially more difficult for black and Hispanic students. Noting the complex and somewhat uncertain nature of these statistical approaches, the California State Department of Education suggests a simplified statistical method to identify especially difficult items.²²⁴ For example, this method could isolate (1) test items which reflect concepts either nonexistent within or unique to one culture, (2) items which include words that have different meanings for different groups, and (3) items with content that reflect experiences more familiar to one

223. The Texas Study, *supra* note 10, is culturally biased almost by definition, since the University of Texas researchers defined incompetent adults as those whose level of mastery of competency objectives is associated with "[i]nadequate income of poverty level or less," "[i]nadequate education of eight years of school or fewer," and "[u]nemployment or low job status." *Id.* at 5. This definition, for example, would include mothers receiving welfare from the Aid to Families with Dependent Children Program, regardless of their actual level of literacy or functional competence. The researchers admit that functional competency is "a construct which is meaningful only in a specific societal context. . . . [It] is culture-bound." *Id.* at 2. The APL test thus does not measure an individual's competence in functioning in that part of society in which he or she lives every day, but instead attempts to measure a person's competency by the test designer's conceptions of what is required for successful functioning in middle-class America. The ability to survive in a ghetto, for example, is not measured by the test, and, therefore, the test results exaggerate the extent of functional incompetence among blacks and other minorities.

224. See California State Dep't of Education, Technical Assistance Guide for Proficiency Assessment app. M (1977) [hereinafter cited as Technical Assistance Guide]. Compare California's approach with the regression and other models discussed in Linn, *supra* note 216.

group than another. The Department defines these items as biased, and recommends either revising or deleting the items. If all or most items measuring a particular skill prove especially difficult, another approach would be to provide special instruction in these skill areas.

Any of these approaches—revision, deletion or special instruction—may resolve the problem, but in some cases it will not. Special instruction may not be sufficient in some cases to overcome minority disadvantage in these areas. Revision or deletion may not be practical if minority students score uniformly lower on most or all items on the test.

At this point some would seek to avoid responsibility with a genetic explanation.²²⁵ Schools would probably contend that the differential test results are not caused by inadequate educational programs, but by the students' poor attitude and/or low socio-economic status.²²⁶ Student attitude is certainly a factor in test performance, but the school may share some responsibility for poor attitude,²²⁷ and the causal relation and interaction between attitude and school is probably as complex as that inherent in the nature-nurture controversy. The most common explanation for poor test performance is low socio-economic status. One report observes: "Undoubtedly factors associated with low socio-economic status contribute to the low test performance of children from these backgrounds. However, very seldom if ever are these socially related factors taken into account by those who develop and generate mass educational programs."²²⁸

The ultimate question raised by any discussion of cultural bias in competency testing is the extent of the school's responsibility, if any, to overcome the disadvantage that many children bring with them when they first enroll in the public school system. This disadvantage is usually the effect of low socio-economic status and what Justice Powell in the *Bakke* case more generally called "societal discrimination".²²⁹

225. See generally *The IQ Controversy* (N. Block & G. Dworkin eds. 1976) (reprinting articles by Jensen, Herrnstein, Jencks, Kamin, and others on whether intelligence has a substantial genetic component).

226. See, e.g., *Larry P. v. Riles*, 343 F. Supp. 1306, 1310-11 (N.D. Cal. 1972), *aff'd per curiam*, 502 F.2d 963 (9th Cir. 1974) (defendant's argument that the disproportionate number of black children in EMR classes simply reflects poor infant care, nutrition, and other environmental factors).

227. For example, teacher expectation can affect student performance and create a "self-fulfilling prophecy." For a dramatic illustration of this point, see R. Rosenthal & L. Jacobsen, *Pygmalion in the Classroom* (1968). Teachers were informed that tests indicated that certain pupils would do well and others poorly; this in fact proved to be the case even though the pupils had been chosen at random. For some of the literature stimulated by the Rosenthal and Jacobsen study, see J. Elasoﬀ & R. Snow, *Pygmalion Reconsidered* (1971); Rist, *Student Social Class and Teachers Expectations: The Self-Fulfilling Prophecy in Ghetto Education*, 40 Harv. Educ. Rev. 411 (1970).

228. *Impact*, *supra* note 25, at 35.

229. 438 U.S. at 307-10.

Equal educational outcomes, or at least a lack of culturally disproportionate outcomes,²³⁰ may seem utopian to many, but that goal does not seem visionary with respect to the minimum skills and knowledge measured by most competency tests.

In any case, it seems clear that competency tests which are culturally biased against racial and linguistic minorities will be subject to legal challenge.²³¹ The unresolved issue is defining "cultural bias". Will the term be limited to obvious forms of bias such as items containing offensive terms and stereotypes? Will it extend to societal discrimination? Or will the line be drawn somewhere in between?

VI. HANDICAPPED STUDENTS

Many of the issues discussed above regarding test content, phase-in periods, instructional validity, and racial and linguistic discrimination apply to handicapped as well as non-handicapped students, but special concerns are raised by competency testing programs for handicapped students.²³²

A. Exemption

Some policymakers may favor total exemption of handicapped students from competency testing programs because of the special difficulty of designing fair assessment procedures for handicapped students, and also because of the emphasis on Individualized Education Programs (IEP) and the annual evaluations of the IEP required by the federal Education for All Handicapped Children Act of 1975.²³³ If handicapped students are exempted from competency testing programs, they should be given the option of taking the test. Failure to provide such an option may deny such students the benefits of an educational program in violation of section 504 of the Rehabilitation Act of 1973.²³⁴ If policymakers decide not to exempt handicapped

230. Equal educational outcomes imply that all students reach a certain educational level. Lack of culturally disproportionate outcomes is a less ambitious goal because it would allow for some students to fall below the specified educational level as long as proportionate numbers of students of all cultures and races were affected. The "effect standard," *see* note 203 *supra* and accompanying text, uses disproportionate outcomes as the trigger for legal analysis.

231. *See* *Morgan v. Kerrigan*, 401 F. Supp. 216, 234 (D. Mass. 1975) (nondiscriminatory curriculum ordered as part of school desegregation plans), *aff'd*, 530 F.2d 401, *cert. denied*, 426 U.S. 935 (1976); *United States v. Texas*, 330 F. Supp. 235, 249 (E.D. Tex.) (same), *modified*, 447 F.2d 441 (5th Cir. 1971). Under Title VI, the curriculum as well as other parts of the school program should not be designed in such a way as to exclude minority children from the benefits of participating in federally funded educational programs. 42 U.S.C. § 2000d (1976).

232. This discussion is based upon McClung & Pullin, *supra* note 5.

233. 20 U.S.C. §§ 1401-1461 (1976). The statute's implementing regulations are codified at 45 C.F.R. § 121a.1-754 (1978).

234. 29 U.S.C. § 794 (1976). The statute's implementing regulations are codified at 45 C.F.R. §§ 80.1-13, 81.1-131, 84.1-99 (1978).

children from the competency testing program, they should consider the issues summarized in this Part before deciding upon the exact nature of the requirement.

B. *Differential Diplomas and Standards*

In assessing the fairness and legality of competency testing programs for handicapped students, the two most fundamental questions are (1) whether there should be differential diplomas for handicapped and nonhandicapped students, and (2) whether there should be differential standards for handicapped and nonhandicapped students.

A *differential diploma* for handicapped students may be defined as a diploma that is distinguishable in any way from that awarded to nonhandicapped students who pass the competency test, for example, in the color, shape, or wording of the diploma. *Differential standards* for handicapped students may be defined as standards that are different from (usually less stringent than) the standards that nonhandicapped students are required to meet. The basic rationale for differential standards is that they are designed to meet the special needs and capabilities of handicapped students.

The three general approaches likely to be considered by policymakers are (1) same diploma and same standards, (2) same diploma and differential standards, and (3) differential diploma and differential standards. These three approaches could be applied with respect to all handicapped students, or handicapped students on an individual basis. For reasons discussed in Part VI(D) and (E), individualized determinations are recommended.

C. *Differential Assessment Procedures*

Another important decision for policy-makers is whether differential assessment procedures should be adopted for handicapped students. This could take the form of modifying the paper-and-pencil test given to regular students, and/or developing methods of assessment that do not require a paper-and-pencil test.²³⁵

The modified paper-and-pencil test is used in Florida where blind and visually impaired students may take Braille or large-print versions of the statewide test given to nonhandicapped students. Modifications of the test instrument, however, are not made for children with other types of handicaps, and for some handicapped children this will raise serious discrimination and equal protection questions under state and federal law.²³⁶

235. Direct or simulated performance measures are among the possibilities. See notes 263-64 *infra* and accompanying text.

236. Various bills to amend the Florida statute to account for these problems have been introduced in the Florida legislature. No amendment has been enacted, however, as of this writing.

Similar legal questions would also be raised for a state that makes no special provisions of this kind for handicapped students, and yet denies diplomas to handicapped students who do not pass the test. The argument would probably be that an unmodified test instrument discriminates against handicapped students, especially those children with sensory or motor problems, on the basis of their handicapped conditions.²³⁷ This argument will be stronger where the modifications are relatively easy to make.

The use of alternatives to a paper-and-pencil test is illustrated by the California statute which provides that local districts may adopt differential assessment procedures for handicapped students.²³⁸ The California State Department of Education emphasizes "that assessment of student performance may be based on multiple criteria, not just a test score."²³⁹

D. *Other Legal Implications*

Although the federal Constitution, statutes, and regulations do not provide clear guidance as of this writing, legal considerations may preclude any uniform approach applied indiscriminately to all handicapped students. What is fair for one handicapped student may be unfair for another because individual circumstances vary so greatly. Some handicapped students (for instance, a student whose only handicap is a speech impairment), who need to be treated like nonhandicapped students but are not, will probably claim violations of the equal protection clause, section 504, and the federal statutory presumption that handicapped students should be integrated into the regular educational program to the maximum extent appropriate.²⁴⁰ Other handicapped students (for example, many, but not all, multi-handicapped persons), who need to be treated differently from nonhandicapped students but are not provided with differential standard and/or assessment procedures, will probably claim violations of the equal protection clause, section 504, and the federal statutory presumption that

237. See generally 45 C.F.R. §§ 84.35(b)(3), 121a.532(c) (1978) (regarding test instruments used for placement purposes for handicapped persons).

238. Cal. Educ. Code § 51215 (West Supp. 1979) provides: "Differential standards and assessment procedures may be adopted for pupils who: (a) Are enrolled in special education programs . . . and (b) Have diagnosed learning handicaps or disabilities that would preclude them from attaining the district's regular proficiency standards with appropriate educational services and support.

"Any such differential standards shall be included in the individualized education program developed for the pupil" For an explanation of this law and discussion of various issues involved in setting differential standards, see Technical Assistance Guide, *supra* note 224, app. L.

239. Technical Assistance Guide, *supra* note 224, at III-15.

240. See 20 U.S.C. § 1412(5) (1976); 45 C.F.R. § 84.34(a) (1977).

to provide equal treatment to persons with unequal needs is unfair in some circumstances.²⁴¹

E. *Individualized Determinations*

Legal and policy considerations suggest that decisions about the nature and extent of participation of handicapped students in any competency testing program should be made on an individual basis. For some handicapped children the fairest approach would be to use the same test, standards, and assessment procedures as those used for nonhandicapped students. For other handicapped children, the fairest approach would entail minor modifications. For many children with severe handicaps, the fairest approach would be completely different tests, standards, and assessment procedures, or even total exemption from the competency test requirement.

Individual decisions of this kind could be made a part of the process of developing the IEP mentioned in Part VI(A). Policymakers in some states, including Missouri,²⁴² Massachusetts,²⁴³ and California,²⁴⁴ are recommending or requiring individual decisions about competency programs for handicapped children based upon the IEP. In California the State Department of Education recommends, among other things, that student proficiency standards for handicapped students be set individually rather than for the group as a whole; that the committee with the responsibility to develop an individual's plan should describe the performance standards in basic skills for which the student may receive a diploma; and that the committee "should prescribe attainable standards which enhance learning."²⁴⁵

VII. RECOMMENDED APPROACHES

The preceding parts of this Article have discussed a number of legal and educational implications of requiring students to pass a competency test as a condition to receiving a high school diploma. For some policymakers, the problems inherent in trying to develop a fair and legal competency testing program will persuade them to forego an

241. In comments to the final regulations under § 504 of the Rehabilitation Act, 29 U.S.C. § 794 (1976), HEW notes that "different or special treatment of handicapped persons, because of their handicaps, may be necessary in a number of contexts in order to ensure equal opportunity." 42 Fed. Reg. 22676 (1977). The regulations require providing aid, benefits, or services when necessary to afford the handicapped effective opportunities. 45 C.F.R. § 84.4(b)(iv)(1978).

242. See *Minimum Competency Testing and The Handicapped: Missouri's Plan*, 59 Phi Delta Kappan 367 (1978).

243. See Proposed Regulations, *supra* note 29, § 40.10(2).

244. See note 239 *supra*.

245. Technical Assistance Guide, *supra* note 224, at III-15.

approach based upon diploma denial, and to explore some nonpunitive, educational alternatives.²⁴⁶ Other policymakers, however, while being concerned about the legal and educational problems, may nevertheless decide to proceed with a diploma-based competency testing program. With these policymakers in mind, this Article will suggest some approaches to competency testing that are designed to minimize the educational and legal problems to the extent possible.

A. *Community Participation*

The decision to implement a competency testing program raises several questions. Should the testing program, for example, be designed to measure only the basic proficiency skills such as reading, writing, and computation, or should it go beyond this by measuring a student's ability to apply these skills in adult-life role activities such as those of a consumer, producer, and citizen? Should satisfactory performance on the test, whatever it measures, be a minimum standard to be used in conjunction with other criteria, or should it be the exclusive criterion of satisfactory performance resulting in a high school diploma regardless of age or course credits?

The author's opinion is that single-criterion evaluation of students, teachers, and public education sells public education short. For one thing, the state of the art is not sufficiently developed to warrant such exclusive reliance on competency-based evaluations. More importantly, the primary goals of public education are or should be broader than those reflected by minimal competencies, and students, teachers, and public education generally should not be evaluated exclusively by these narrow measures. Even given a strong commitment to a broader view of public education, there is always the danger that the minimal standards will become maximums rather than minimums. As in other areas such as drivers' licenses and bar examinations, however, the need for establishing minimum standards has been demonstrated, and safeguards can be developed to ensure that broader goals are not bypassed.

These issues are subject to considerable difference of opinion, as illustrated by early exit programs that allow students, regardless of age, to graduate from high school upon passing a minimum competency test. Competency testing programs obviously involve implicit or explicit decisions about performance objectives and educational goals,

246. For example, the National Academy Report, after criticizing competency testing as a prerequisite to a high school diploma, states: "However, the Panel is in agreement that a series of standardized tests at the lower grade levels used for diagnosing individual student weaknesses, pinpointing remediation needs, and building public pressures if school-wide performances in basic skills continue over time to be consistently low, could be positive influences on student learning." National Academy Report, *supra* note 3, at 9.

and these in turn have important implications not only for curriculum and instruction but also for other school practices such as grouping and discipline.²⁴⁷ Given the crucial importance of these decisions, a model program should provide for representative community-based participation in the decisionmaking process.²⁴⁸

There may also be a legal rationale for community involvement since the standard-setting process is essentially an arbitrary and judgmental one. Criterion-referenced or objective-referenced measures may appear objective and scientific, but in fact they are based on subjective judgments and subject to the charge of arbitrariness, as illustrated by Glass's critique²⁴⁹ of the Florida competency test²⁵⁰ and also by the corresponding claim by plaintiffs in *Debra P. v. Turlington*.²⁵¹ Determining the cut-off score for a norm-referenced test is also, of course, somewhat arbitrary. This does not mean that the arbitrariness is legally vulnerable arbitrariness. Courts are no strangers to arbitrary standards; in fact, they set many themselves to aid resolution of difficult issues. Although most courts would be reluctant to conclude that a cut-off score on a competency test is illegally arbitrary,²⁵² the kind of arbitrariness inherent in both criterion-referenced and norm-referenced tests may focus judicial attention on the process of setting the educational goals, performance objectives, and cut-off scores. A court might invalidate a test if the standards have not been "legitimized" by community participation or some other representative process.

Whether based on legal or political considerations, community-based participation is required by many competency testing statutes.²⁵³ But community participation in turn is likely to aggravate another

247. See, e.g., Spady, *supra* note 1, at 9-14; Spady & Mitchell, *supra* note 160, at 9-15.

248. Compare Shepard, *Setting Standards and Living With Them*, 18 Fla. J. Educ. Research 28, 28-32 (1976) (recommending that standard setting be an iterative process involving various audiences).

249. Glass, *supra* note 27.

250. See *As Always*, *supra* note 84; notes 83-85 *supra* and accompanying text.

251. "The state defendants have arbitrarily determined what scores will be sufficient to pass the test." Complaint, *supra* note 63, at ¶ 46.

252. Plaintiffs, nevertheless, will probably argue that the cut-off score for classifying persons as either competent or incompetent is so arbitrary as to be unconstitutional. Their general argument is articulated by Judge Davis: "When a program talks about labeling someone as a particular type and such a label could remain with him for the remainder of his life, the margin of error must be almost nil." *Merriken v. Cressman*, 364 F. Supp. 913, 920 (E.D. Pa. 1973) They will probably produce expert testimony questioning the educational and psychometric validity of making important judgments about an individual on the basis of a single test.

253. Cal. Educ. Code § 51215 (West Supp. 1979), for example, provides: "Standards of proficiency shall be adopted by the governing board with the active involvement of parents broadly reflective of the socioeconomic composition of the district, administrators, teachers, counselors, and, with respect to standards in secondary schools, pupils." See also Proposed Regulations, *supra* note 29, § 40.02(2).

problem—the kind of mismatch between test and instruction mentioned in Part IV. By including broad participation in the standard-setting process, states and school districts are likely to end up including some new or at least redefined educational goals and performance objectives. Thus there is always a potential conflict between the educational and perhaps legal need for some kind of match between test and instruction and the community's political need to set standards that do not simply rubberstamp the educational status quo. In order to resolve this conflict, states and/or school districts may want to consider the kind of two-phase implementation procedure discussed in the next section.

B. *Two-Phase Implementation Procedure*

In attempting to resolve the questions of legality and fairness raised by this Article, the persons responsible for developing a competency program may want to adopt a two-phase implementation procedure. Phase I would emphasize that the competency standards for the first classes of students affected by the new requirement should be formulated on the basis of the curriculum and instruction to which these students have already been exposed for most of their educational careers in the district's classrooms. In other words, during phase I the focus would not be on which proficiency skills the governing board and community think *should be* taught and emphasized in the district's schools but rather on those proficiency skills that *have been* taught and emphasized previously by the district. In some schools there will be no actual distinction between what should be emphasized and what has been emphasized, but in other schools the distinction will be substantial. The distinction will probably be reflected in the difference between basic literacy and numeracy skills emphasized by most schools and the adult-life skills emphasized by most competency tests.

Thus, phase I standards are essentially past standards—those that have already been emphasized in district classrooms. Students, parents, and teachers should still be notified of phase I standards as soon as they are adopted. Even though the standards will not have changed, both the assessment procedures and the penalty for failure to meet the standards, denial of a diploma, are new; therefore, advance notice is essential.²⁵⁴ Phase I standards should be retained until the new curriculum and instruction, if any, have been implemented and adequate notice is given to students, parents, and teachers.

New curriculum and new standards, if any, are the focus of phase II. Given the severe nature of the penalty to be imposed on students who cannot meet the district's phase I standards, a state or school district might, as a policy matter, want to reevaluate its traditional

254. See notes 144-48 *supra* and accompanying text.

standards and develop new, more appropriate standards. The planning for phase II should begin simultaneously with the implementation of phase I.

The suggested two-phase implementation procedure is based on a number of legal and policy conclusions discussed in this Article. Perhaps the most important of these conclusions are: (1) the need for some kind of match between test and instruction (instructional validity), and (2) the importance of the distinction between measures of basic literacy/numeracy skills and adult-life skills.

For those who agree with these conclusions and supporting rationale, the two-phase implementation procedure will help to ensure fairness for all students and will thereby reduce the risk of legal challenge. This approach by itself, however, cannot resolve all concerns about fairness and legality. The most serious of these concerns is the potential discriminatory effect of competency testing programs on racial and linguistic minorities.

If a state or school district decides to develop a second phase, it might want to consider following the five minimum steps suggested in Part VII(D)(1) as a basic framework for setting new graduation requirements—starting with substantial community involvement and concluding with a careful analysis of the psychometric and instructional validity of the assessment procedures. Some suggestions concerning instructional validity and assessment follow.

C. *Approaches to Instructional Validity and Assessment*

The assessment procedures, whether developed as part of phase I or phase II, should meet professional psychometric standards and have instructional validity. This Section will describe some approaches to instructional validity and consider direct performance measures as an alternative or supplement to the traditional paper-and-pencil test.

In developing or selecting an assessment instrument, steps should be taken to ensure the instructional validity of the instrument.²⁵⁵ Although content validity is a distinguishable concept,²⁵⁶ it is sufficiently related to offer guidance in determining the instructional validity of a

255. The approaches discussed herein to assess the instructional validity of a competency test reflect preliminary research and thoughts. Better approaches might be developed. For example, Laura Wagner of the Stanford Graduate School of Education suggests that the instructional audit developed by Professor Daniel Duke of Stanford University, or some adaptation of the instructional audit, might be useful in establishing the instructional validity of a competency test. See L. Wagner, *The Duke Instructional Audit: An Aid in Curricular Decision-Making* (Nov. 22, 1978) (paper presented to National Council for the Social Studies). Perhaps methods to determine instructional validity, including the above approach, can be developed to generate the necessary information without significantly increasing recordkeeping or infringing upon classroom instruction.

256. See note 155 *supra* and accompanying text.

proposed instrument. One approach is suggested by ETS's study²⁵⁷ of the use of the National Teacher Examination (NTE) for certification purposes in North Carolina. The study examined the content domain of the NTE in relation to the content domain that should be measured if scores are used for initial certification of teachers. The researchers determined that the content domain that the test should measure was the content of teacher education programs offered by North Carolina colleges:

There are four sets of data that are measures of the correspondence between the content of the test and the content of the teacher training program: (1) the percentage of questions classified as content appropriate; (2) the number of content areas identified as omitted from the test; (3) the judgments about relative emphasis on specified subject matter in the test and in the teacher training program; and (4) the overall judgments with respect to whether the test parallels the curriculum. Criteria were established with respect to each of these sets of data to identify the degree of relationship between the test content and the program content. These criteria were applied individually to each test in the Common Examinations and each of the Area Examinations, and the interrelationships of criteria were also considered.²⁵⁸

This kind of content validation of the NTE was also conducted by the ETS with respect to teacher education programs in South Carolina, and secured judicial acceptance in *United States v. South Carolina*.²⁵⁹ This approach could be adapted to measure instructional validity by relating test content to instruction in specific schools rather than to curriculum on a statewide basis.²⁶⁰ Furthermore, a sound approach would support judgments about instructional validity with evidence of actual instruction.

Schools will have to play a key role in the development of the assessment instrument even if, and perhaps especially if, they contract with an independent test publisher. Popham has noted that competency testing poses problems for test publishers because the tests will have to be based upon curriculum to a far greater extent than the traditional standardized tests, thereby making the publishers' instruments usable only by districts with very similar curricula.²⁶¹ This problem will be especially acute in states where each district is responsible for developing its own standards.

257. Educational Testing Service, Report on a Study of the Use of the National Teacher Examinations by the State of North Carolina (1976).

258. *Id.* at 159.

259. *United States v. South Carolina*, 15 Fair Empl. Prac. Cas 1196 (D.S.C. 1977), *aff'd mem. sub nom.* National Educ. Ass'n v. South Carolina, 434 U.S. 1026 (1978); see notes 168-71 *supra* and accompanying text.

260. Readers interested in this approach will want to refer to the complete report for details on how these determinations were made and what degree of match between test and program was considered sufficient. See Educational Testing Service, Report on a Study of the Use of the National Teacher Examination by the State of North Carolina 5 (1976) (chart and related explanation). The reference to the report herein is for illustrative purposes only and is not intended as an endorsement of the approach described in the report.

261. Popham, *Customized Criterion-Referenced Tests*, 34 Educ. Leadership 258, 258-59 (1977).

One way in which test publishers are attempting to solve this problem is by creating "item pools." Districts that have specifications based on their performance objectives can match their specifications with those for which items and exercises have already been prepared by the test publisher. Whatever approach is used in test development or test selection, the district will need to make important determinations about whether the proposed instrument is sufficiently related to curriculum and instruction. The following advice that Melville offers school districts in selecting an achievement test is generally applicable to competency tests:

[T]he most important elements in good test selection [are]: How well does the test measure what it was selected to measure? Are the things being measured important? Is there a proper balance of areas covered? Are all the important areas to be taught included? *No one*—salesman, neighboring school colleague, learned reviewer—can answer these questions. Only the classroom teacher and his department colleagues can supply the answers, and they can only do so by looking carefully at the actual test questions one by one.

The task of making a careful examination of a test cannot be simplified very much. It is a time-consuming job. It can be more effectively accomplished, however, if approached systematically. One approach, applicable primarily to the selection of an achievement test, follows.

The objective of the test analysis is to ascertain (1) what kind of materials are included in the test and (2) the relevance of these materials to the teaching program. In order to appraise test items in a meaningful way, one should ask of each item: What particular skill, understanding, type of material, or subject matter does this item seem to be covering? Since judgments regarding the relevance of an item are necessarily subjective, a rough point-value scale for judging relevance should suffice. Items considered inconsequential or trivial would be rated 0 to 1; those closely related to local objectives and student capabilities would be rated 4 or 5. A work sheet ensures that the analysis is done systematically. Four column headings are needed; item number, item content, skills involved, and relevance of the item.²⁶²

The approach suggested by Melville could be adapted to make determinations of test relevance to curriculum and instruction. The district should ensure a sufficient match between curriculum and instruction by eliminating tests or test items that are not sufficiently matched with curriculum and instruction.

A sound approach to instructional validity would not rely exclusively on the subjective judgments of a panel of administrators or teachers about whether specific test items were taught in the district's classrooms. These judgments should be supplemented with, and supported by, comparisons of test items and evidence of a district's curriculum and instruction. Such comparisons would include evidence such as curriculum guides, textbooks and materials used in the classroom, lesson plans, student homework samples, and perhaps inter-

262. S. Melville, *Selecting an Achievement Test* (four-page summary of the script from a film with the same title, available from the Cooperative Test Division, Educational Testing Service, Princeton, N.J.).

views with a representative group of students and teachers. A school district concerned about the instructional validity of its assessment procedure might want to arrange for an outside agency to examine such evidence and prepare a report of its findings and conclusions prior to final approval of the assessment procedure.

Madaus and Airasian write that many of the minimal cognitive competencies for graduation involve application of basic literacy and numeracy skills to real-life situations, for example, checking the accuracy of bills and sales slips, using the public library, using the town and state offices. "These competencies are most validly measured by the most direct means possible, situational or performance examinations which determine if the student can actually perform the behaviors."²⁶³ Although they recognize that direct measurement is often costly and time consuming, the two authors conclude that "indirect paper-and-pencil tests, measuring knowledge about the competency areas, are not enough. Any indirect, or surrogate, measurement must be validated against direct performance measures."²⁶⁴ Serious questions about fairness and legality would be raised if a student could show that he or she was denied a diploma on the basis of performance skills that he or she could demonstrate by direct assessment but not by the indirect paper-and-pencil method. This problem also suggests that if a student cannot pass a paper-and-pencil competency test, perhaps that student should be given a direct performance test to be sure that he or she does not have the requisite skills before denying him or her a diploma.

Field testing of the proposed test can serve several important purposes, including determination of its likely effects on blacks, Hispanics, and the student population as a whole. If substantial or disproportionate numbers of students in any category cannot pass the proposed test, the state or school district may want to take another look at the adequacy of the match between the test and instruction in its classrooms and at the reasonableness of its proposed cut-off score. Where limited and non-English-speaking students have difficulty with the field test, the test could be given in the student's native language in order to determine whether the student's difficulty on particular test items is with language or the skills measured by the items. Such assessments can be used to help provide low-scoring students with appropriate remedial instruction before the final assessment.

D. *Proposals for Competency Testing Programs*

1. Basic Framework

Before new graduation standards are used as a basis for denying any student a diploma, a sound and careful competency testing program

263. Madaus & Airasian, *supra* note 15, at 86.

264. *Id.*

would *at a minimum* include the following as its basic framework:

1. Provide for substantial parental and other community involvement in developing and adopting educational goals, performance objectives, and assessment procedures.
2. Provide sufficiently detailed advance notice of the new educational goals, performance objectives, and assessment procedures to students, parents, and teachers.
3. Make necessary changes in curriculum to reflect new educational goals and performance objectives.
4. Provide for a sufficient phase-in period of the new curriculum and instruction. Sufficient phase-in may take years, depending upon the degree of curricular/instructional change.
5. Take steps to ensure that the assessment procedure meets professional psychometric standards and has instructional validity, as defined in Part IV(A).

2. Model Program Provisions

Policymakers may also want to consider the following model program provisions in attempting to design a fair and equitable program.

A committee or committees of representative community, school, and professional persons should be formed to review the proposed test in order to eliminate any items with inappropriate content, including (1) coerced belief, (2) invasion of privacy, (3) unteachable or unmeasurable content, (4) content that is not sufficiently matched with a school's curriculum and instruction, and (5) content biased on grounds of race, language, culture, national origin, sex, handicap, etc. Particular emphasis should be given to legal representation for a review of the first two categories, to psychometric and classroom-teacher representation for the last category.

The proposed assessment procedure should be pretested, and the test results used to help determine (1) a reasonable cut-off score or procedure, and (2) whether the test may be biased against any racial, linguistic, or other group. A disproportionate failure rate on the test generally or on specific items for any particular group is an indication of possible bias. Statistical approaches can be used to determine which items are especially difficult for various groups.²⁶⁵

A model program should provide for multiple learning, evaluation, and remedial opportunities. Students who will have difficulty passing the test, as identified by the pretesting, should be given sufficient instruction and/or remedial instruction to help them pass the test. Provision should be made to insure that students who do not pass the test are not tracked in all courses just because remedial instruction is

265. See note 224 *supra* and accompanying text. *But see* Linn, *supra* note 216 (limitations of statistical approaches).

necessary, especially if remedial classes would involve racial segregation or resegregation of students.

Where adequate education and/or remediation for some students is not possible in the time available before the final test, such students should not be required to pass the test as a prerequisite to receiving a regular high school diploma. Exemption of black students from this requirement may be appropriate, for example, where the school district is under a court desegregation order or agreement, or where there has been some other finding of prior discrimination by the school against these students. Another example would be school districts that have initiated, but have not had time to implement adequately bilingual programs for limited and non-English speaking students.

A model program should provide for differential standards and/or assessment procedures for handicapped students. These standards and procedures should be determined on an individual basis by the team that formulates an individualized educational program for the student pursuant to the Education for All Handicapped Children Act of 1975.²⁶⁶

A low score on a competency test should be used as the trigger for a more careful evaluation of the student's performance rather than as the final determination of failure. The program should utilize a committee to determine by other methods whether the student possesses the requisite knowledge and skills, especially if the student has achieved a near-pass score.²⁶⁷ Simulated or direct performance measures could be used where appropriate to make sure that the indirect paper-and-pencil measure is accurate before denying any student a diploma. Former students who failed the test should be given an option of earning the high school diploma by further remedial education and test opportunities at any later point in their lives.

Since learning is a two-way street, a model program should provide for students and educators to share responsibility for performance rather than place the full burden on the students. For example, if passing a competency test is one criterion for a student's earning a high school diploma, perhaps a teacher's success in helping students pass the competency test should be one criterion in the various forms of teacher evaluation, for example, tenure decisions. Such evaluation is feasible if students are tested at the beginning and end of each year in each course, and allowance is made for mitigating circumstances such as student absences and transfers.

266. 20 U.S.C. §§ 1401-1461 (1976); see notes 240-45 *supra* and accompanying text.

267. Instead of using a single cut-off score to determine minimally acceptable performance on a competency test, some educators suggest a cut-off "band," whereby students with near-pass scores within the band are evaluated by other means before making a final determination about performance.

Finally, administrators should be responsible for making the necessary changes in school practices, and the public should be responsible for financing such changes. If these corresponding responsibilities for teachers, administrators, and the public seem unreasonable or unfair, then perhaps the reasonableness and fairness of denying diplomas to students based upon the results of a competency test should be reconsidered. Developing a workable model of shared responsibility will obviously be more difficult than placing the entire burden on students, but it will also be fairer.

CONCLUSION

This Article has identified a number of legal and educational problems raised by most programs that make passing a competency test a prerequisite to receiving a regular high school diploma. The potential of these programs for unfairness, combined with the severe nature of the penalty—which could in effect relegate a student to second-class citizenship—makes legal challenge likely and special care by state legislatures and school districts essential.

Since litigation is costly and difficult for all parties, states and school districts that are committed or inclined to a diploma-based competency program may want to seek independent legal opinion about the legality of their proposed program before it ends up in court. In the absence of more specific materials, this Article can be used to anticipate plaintiffs' arguments. The next step is to evaluate the strength of these arguments in light of the exact competency testing program proposed, and the exact factual situation presented by the state or school district. If independent legal opinion indicates serious problems, the state or school district may want to modify the proposed program to meet the legal concerns. States and school districts that are not already committed to a diploma-based competency program may want to consider alternative approaches that raise fewer legal and educational problems.

The ultimate arbiters of the legality of competency testing programs are, of course, the courts. As various states and districts begin to withhold diplomas on the basis of competency tests, and legal challenges are heard by the courts, a judicial pattern and predictable case law on competency testing will eventually take shape. This case law should be read with care as it will set the basic parameters of educational policymaking.

It should be emphasized, however, that whatever the shape of these legal parameters, they will reflect only the minimum standards essential to fairness under our legal system. Policymakers must meet, but

are not limited to, these minimum standards in pursuing the goal of educational equity for students. Some of the legal arguments, even if unsuccessful in the courts, suggest standards of fair play. For example, even if the eventual case law on competency testing does not require some kind of match between what is measured by the test and what has been taught in the school, most persons would probably agree that this is an essential component of any fair competency testing program where high school diplomas are at stake. Similarly, prior notice to students and gradual phase-in periods will probably strike most persons as important even if the courts decide that they are not legally required.

The discussion set forth at the outset of this Article indicates substantial disagreement about the advisability of competency testing as a prerequisite to a high school diploma. Whatever one's personal views about the competency testing movement, most readers would probably agree that, except in unusual cases, functional illiteracy after twelve years of public education is simply unacceptable. Perhaps this concern can be mobilized to develop more equitable programs and to provide an adequate education for all children.