

Spring 1991

## Diploma Sanctions Tests Revisited: New Problems from Old Solutions

S.E. Phillips

Follow this and additional works at: <https://scholarcommons.sc.edu/jled>



Part of the [Law Commons](#)

---

### Recommended Citation

S. E. Phillips, Diploma Sanction Tests Revisited: New Problems from Old Solutions, 20 J.L. & EDUC. 175 (1991).

This Article is brought to you by the Law Reviews and Journals at Scholar Commons. It has been accepted for inclusion in The Journal of Law and Education by an authorized editor of Scholar Commons. For more information, please contact [digres@mailbox.sc.edu](mailto:digres@mailbox.sc.edu).

# Diploma Sanction Tests Revisited: New Problems From Old Solutions

S. E. PHILLIPS\*

Lake Wobegon is a mythical town in Minnesota where "all the women are strong, all the men are good-looking, and all the children are above average."

—Garrison Keillor, *A Prairie Home Companion*<sup>1</sup>

We achieve success the old-fashioned way: we earn it.

—John Houseman<sup>2</sup>

## I. Introduction: Defining The Problem

Consider the following situation that I have posed to my graduate education classes. Suppose you are interviewing for a cashier job in a drug store and have a choice of two evaluation methods: (1) you can take a test of basic numerical skills (e.g., addition of decimals; determination of correct change) OR (2) you can be interviewed by Archie Bunker. Which would you choose?

My students are typically divided in their responses to this query, but most minority students choose the second option. They seem to believe (often quite strongly) that the test will be biased against them but that they stand a chance of being hired if they have the opportunity to talk to the boss. What they seem to be overlooking is that interviews are purely subjective and the results may be based on criteria that are capricious and idiosyncratic to the interviewer.<sup>3</sup> Therefore, if the interviewer is prejudiced against the (short, fat, female, minority, etc.) applicant, the interviewer can certainly invent some basis for finding the applicant unsuitable.<sup>4</sup> On the other hand, if the ground rules for hiring are based on

---

\* BA in mathematics, University of Iowa; MA, University of Iowa; PhD, University of Iowa; JD, Thomas M. Cooley Law School. The author currently is Associate Professor of Education at Michigan State University and conducts research on issues in education law.

1. Haney, *Editorial*, 7(2) EDUC. MEASUREMENT: ISSUES & PRAC. 4 (1988) (describing Garrison Keillor, *A Prairie Home Companion* (American Public Radio (1975-1988)) (prologue to regular broadcast).

2. Paraphrase of John Houseman commercial for Smith-Barney in which he says, "We make money the old-fashioned way. We earn it." See also Samuelson, *The Binge is Over*, NEWSWEEK, July 10, 1989 at 35 ("We will not have to get our prosperity the old-fashioned way. We will have to earn it.").

3. THORNDIKE & HAGEN, MEASUREMENT AND EVALUATION IN PSYCHOLOGY AND EDUCATION, 397, 400, 458 (4th ed. 1977).

4. *Id.*

a numerical skills test, the person performing best on the numerical basic skills test gets the job, regardless of physical appearance or other irrelevant characteristics.<sup>5</sup>

Real employers are likely to use a combination of both methods in selecting employees.<sup>6</sup> In fact, employers often rely on the public schools to provide an initial indication of skill level for entry level jobs by requiring applicants to have a high school diploma.<sup>7</sup> However, in recent years, employers have found that many job applicants with high school diplomas lack basic reading, writing and mathematics skills.<sup>8</sup> Estimates indicate that as many as 20% of all 17-year-olds may be functionally illiterate and anumerate; they also tend to be generally unemployable, subsisting primarily on welfare or crime.<sup>9</sup> The result is that many high school graduates who get a job by having a high school diploma soon lose it because they are unable to perform adequately.<sup>10</sup>

To address this problem, several states have implemented statewide competency tests which all students must pass in order to receive a high school diploma.<sup>11</sup> According to the 1988 U.S. Department of Education Wall Chart, twenty-two states, plus the District of Columbia, require students to pass a competency test in order to receive a high school diploma.<sup>12</sup> Public opinion supports these competency tests because they transfer accountability to the public schools.<sup>13</sup> However, these tests have been severely criticized because of their disparate impact on minorities.<sup>14</sup>

---

5. American Psychological Association, *AERA-APA-NCME Standards for Educational & Psychological Testing*, 1 (1985).

6. See generally W. J. POPHAM, *EDUCATIONAL EVALUATION* 82 (2nd ed. 1988).

7. See, e.g., Note, *Education and the Law: Functional Literacy Program — A Matter of Timing*, 10 STETSON L. REV. 125, 128-29 (1980) (reporting that the State of Florida hires only about 10% of its employees without high school diplomas and employs them in menial jobs); Note, *High School Exit Tests and the Constitution: Debra P. v. Turlington*, 41 OHIO ST. L.J. 1113, 1123-24 (1980).

8. NOTE, *High School Exit Tests and the Constitution: Debra P. v. Turlington*, 41 OHIO ST. L.J. 1113, 1115 (1980).

9. Lerner, *Vouchers For Literacy: Second Chance Legislation*, 63 PHI DELTA KAPPAN 252 (1981).

10. Note, *Constitutional Law — Discrimination — Passage of Functional Literacy Examination as Prerequisite to Receipt of High School Diploma Held Unconstitutional*, 10 CUMB. L. REV. 863, 879, n. 69 (1980).

11. Miller, "Wall Chart" Data Indicate Plateau In Reform Drive, *EDUC. WK.*, May 10, 1989, at 1, 16-17. For additional educational reform statistics, see Elliott & Hall, *Indicators of Performance: Measuring the Educators*, 4(2) *EDUC. MEASUREMENT: ISSUES & PRAC.* 6 (1985).

12. *Id.*

13. Mehrens & Kaminski, *Methods for Improving Standardized Test Scores: Fruitful, Fruitless, or Fraudulent?*, 8(1) *EDUC. MEASUREMENT: ISSUES & PRAC.* 14 (1989); Calfee, *Standards, Evidence and Equity: Implications of the 1983 Debra P. Decision*, 2(4) *EDUC. MEASUREMENT: ISSUES & PRAC.* 11 (1983); Elliott & Hall, *Indicators of Performance: Measuring the Educators*, 4(2) *EDUC. MEASUREMENT: ISSUES & PRAC.* 6, 7 (1985).

14. Williams, *Stimulus/Response: Scientific Racism & IQ — The Silent Mugging of the Black Community*, 7(12) *PSYCHOLOGY TODAY* 32 (1974).

Specifically, critics have challenged the testing requirement impact on minorities, applying equal protection and procedural due process arguments.<sup>15</sup> *Debra P. v. Turlington* is the landmark case making an equal protection and due process challenge to Florida's competency tests.<sup>16</sup>

## II. Background

### A. *Debra P. v. Turlington*

*Debra P.* was a class action suit brought by black students who failed the Florida *Functional Literacy Examination (SSAT II)*.<sup>17</sup> Passage of the SSAT II<sup>18</sup> was a statutory requirement for receipt of a high school diploma in Florida. A 1978 legislative amendment to the Educational Accountability Act of 1976 made the requirement effective for the 1979 graduating class.<sup>19</sup> The statute also required completion of a locally-specified minimum number of course credits.<sup>20</sup> Students who completed the required number of credits but failed the SSAT II were to be awarded a certificate of completion instead of a high school diploma.<sup>21</sup>

At the time the initial suit was filed, the SSAT II had been administered three times: Fall 1977, Fall 1978, and, Spring 1979.<sup>22</sup> In all three administrations, the percentage of black students failing the SSAT II was greater than the percentage of white students who failed the test.<sup>23</sup>

---

15. See, e.g., *Debra P. v. Turlington*, 474 F. Supp. 244 (M.D. Fla. 1979), *aff'd, in part, rev'd, in part*, 644 F.2d 397 (5th Cir. 1981); *Anderson v. Banks*, 520 F. Supp. 472 (S.D. Ga. 1981), *reh'g* 540 F. Supp. 761 (S.D. Ga. 1982).

16. The *Debra P.* case involved two separate trials, four years apart. The initial case was brought in 1979 in Federal District Court and was appealed, *Debra P. v. Turlington*, 474 F. Supp. 244 (M.D. Fla. 1979), *aff'd, in part, rev'd, in part*, 644 F.2d 397 (5th Cir. 1981). The case was remanded for further proceedings at the expiration of the injunction granted by the court. The second trial occurred in 1983 and was also appealed, *Debra P. v. Turlington*, 564 F. Supp. 177 (M.D. Fla. 1983), *aff'd*, 730 F.2d 1405 (11th Cir. 1984).

17. *Debra P. v. Turlington*, 474 F. Supp. 244 (M.D. Fla. 1979), *modified and remanded*, 644 F.2d 397 (5th Cir. 1981). The name of the test was later changed to the *State Student Assessment Test, Part II*, hereafter referred to as the SSAT II. This was due in part to the negative connotations of the term "functional literacy." *Debra P.*, 474 F. Supp. at 258-59. The Florida Department of Education had defined functional literacy as "the satisfactory application of basic skills in reading, writing and arithmetic, to problems and tasks of a practical nature as encountered in everyday life." *Id.* at 258. For a complete list of the test objectives, see *Id.* at 259, n. 22.

18. The SSAT II was a test of basic mathematics and communications skills applied to everyday life situations. It covered 11 communication and 13 mathematic skills. *Debra P. v. Turlington*, 564 F. Supp. 177, 179-80 (M.D. Fla. 1983).

19. *Debra P. v. Turlington*, 474 F. Supp. 244, 247-48, 265 (M.D. Fla. 1979).

20. *Id.* at 247.

21. *Id.* at 248.

22. *Id.* at 248.

23. *Id.* at 248-49. On the first administration of the SSAT II in the fall of 1977, 78% of the black students failed one or more sections (communications and mathematics) while only 25% of the white

Students who failed an earlier administration of the SSAT II were allowed to retake the test at subsequent administrations.<sup>24</sup> After the third administration in the spring of 1979, 20% of the black high school seniors had still *not* passed the SSAT II, while only 2% of the white seniors had *not* passed.<sup>25</sup> The combination of these statistics and the short preparation time for the 1979 graduating class to be instructed in the tested skills provided the basis for the black students' suit. The students alleged that the testing requirement was: (1) racially biased in violation of the fourteenth amendment equal protection clause,<sup>26</sup> Title VI of the Civil Rights Act of 1964,<sup>27</sup> and the Equal Educational Opportunity Act of 1976;<sup>28</sup> (2) instituted without adequate procedural notice as required by the due process clause of the fourteenth amendment;<sup>29</sup> and, (3) used to resegregate Florida public schools by tracking blacks who failed the SSAT II into remedial classes composed almost entirely of black students which violates the same federal constitutional and statutory provisions listed in the first allegation.<sup>30</sup>

The Federal District Court held that there was *no present* intent to discriminate.<sup>31</sup> However, the court concluded that the effects of past intentional discrimination by Florida were perpetuated by the test and, thus, the equal protection and statutory violations were established.<sup>32</sup>

The court further found that 13 months was *not* adequate notice under a due process standard.<sup>33</sup> The court suggested that four to six years' advance notice of the testing requirement should be given to students so that

---

students failed. The failure rates for the second administration were 74% and 25%, respectively, for black and white students taking one or both sections of the test for the second time. By the third administration, the failure rates for repeat takers had evened out at 35% for black students and 30% for white students. *Id.*

24. *Id.*

25. *Id.* at 249. Thus, if the testing requirement were upheld, 3,466 black seniors and 1,342 white seniors, or a total of approximately 5,300 high school seniors (5.8%) would receive a certificate of completion (if they completed their coursework) instead of a high school diploma. *Id.*

26. U.S. CONST. amend. XIV, § 1.

27. 42 U.S.C. § 2000d (1976).

28. 20 U.S.C. § 1703 (1976).

29. U.S. CONST. amend. XIV, § 1.

30. *Debra P. v. Turlington*, 474 F. Supp. 244, 247 (M.D. Fla. 1979).

31. Note, *Constitutional Law — Discrimination — Passage of Functional Literacy Examination as Prerequisite to Receipt of High School Diploma Held Unconstitutional*, 10 CUMB. L. REV. 863, 873 (1980).

32. *Debra, P.*, 474 F. Supp. at 254-55. See also Note, *supra* note 8, at 1131. ("The perpetuation doctrine posits that ostensibly neutral government acts cannot reinforce past constitutional violations.")

33. *Debra P.*, 474 F. Supp. at 267. See *Mathews v. Eldridge*, 424 U.S. 319 (1976) (required due process procedures to be determined by a balancing test which considers three factors: (1) affected private interest, (2) risk of erroneous deprivation and value of alternative procedures, and (3) government interest).

they might adequately prepare to demonstrate skills that are typically learned in multiple classes over a period of several years.<sup>34</sup> Finally, the District Court concluded that blacks were overrepresented in remedial classes *not* because of an intent to resegregate but because substantially more blacks failed the test.<sup>35</sup>

To remedy the vestiges of discrimination and due process violations, the court enjoined the use of the SSAT II as a diploma sanction until 1983, the year in which all blacks would have completed 12 years of education in a unitary system.<sup>36</sup> The court also ordered that resegregation through remediation was permissible only if the classes “will remedy the present effects of past discrimination through better educational opportunities.”<sup>37</sup>

### ***B. Appeal of Debra P.***

On appeal of the District Court decision, two major issues emerged as crucial to the determination of whether Florida’s testing program violated any federal laws: (1) the effects of segregation on the black students’ achievement; and (2) whether the test had curricular validity.<sup>38</sup> The appeals court affirmed the temporary injunction prohibiting Florida from requiring passage of the SSAT II for a high school diploma until the 1983 graduating class. Further, the Florida appeals court affirmed the lower court’s finding that remediation based on SSAT II results was constitutional and violated *no* federal law. Finally, the court reversed the lower court on the issue of whether the SSAT II had curricular validity.<sup>39</sup> The appeals court remanded the case to determine: (1) the curricular validity of the test; and (2) whether the vestiges issue would be moot in 1983.<sup>40</sup>

## **III. Curricular and Instructional Validity**

### ***A. Definitions***

The *Debra P.* court used ‘curricular validity’ to describe its requirement

---

34. *Debra P.*, 474 F. Supp. at 267.

35. *Id.* at 268. Critics claimed that the test was invalid because black students had *not* been taught the skills tested and, conversely, that blacks were discriminated against more than whites by being placed in remedial classes designed to teach the skills tested. *Id.*

36. *Id.* at 269.

37. *Id.* at 268.

38. *Debra P. v. Turlington*, 644 F.2d 397, 407-08 (5th Cir. 1981). The trial court’s finding the SSAT II valid was held by the appeals court to be clearly erroneous due to the lack of sufficient supporting evidence in the record. *Id.* at 402.

39. *Id.* at 408.

40. *Id.* The court essentially held that prior to 1983, the test created suspect classes based on past racial discrimination which subjected Florida’s test to strict scrutiny. After students were freed from the effects of the prior dual educational system, a “rational basis” standard was applicable. Note, *supra* note 8, at 1130.

that the test be fair; but the court's description of this requirement has been more commonly associated with the term "instructional validity."<sup>41</sup> Curricular validity requires that what is tested be contained in the texts, workbooks, and other instructional materials constituting the curriculum.<sup>42</sup> Instructional validity requires that what is actually taught to students in the classroom correspond with what is tested.<sup>43</sup> Curricular validity is typically measured by the degree of overlap between the test questions and textbook exercises classified on the same taxonomy matrix; instructional validity is quantified by measuring the overlap between the test content and teachers' objectives or classroom observations.<sup>44</sup> If teachers deviate from the published curriculum, a test could have curricular validity but *not* instructional validity.<sup>45</sup> A statewide test should have both curricular and instructional validity, although instructional validity is more important.<sup>46</sup> Instructional validity is a more demanding and costly standard to measure.<sup>47</sup>

The court in the *Debra P.* case implied that the required standard was instructional, rather than curricular, validity. Dr. Robert Linn stated that instructional validity is "a clearly unattainable standard because surely the state could not provide evidence that *every* teacher covered *all* the material in the designated curriculum . . . Practicality will necessitate acceptance of a less stringent requirement."<sup>48</sup>

### B. Curricular/Test Match

Proponents of curricular validity argue that overlap between the curriculum and the test adequately informs students what they must learn to

---

41. Linn, *Curricular Validity: Convincing the Courts That It Was Taught Without Precluding the Possibility of Measuring It*, in *THE COURTS, VALIDITY, AND MINIMUM COMPETENCY TESTING* 115 (G. Madaus ed. 1983).

42. McClung, *Competency Testing Programs: Legal and Educational Issues*, *FORDHAM L. REV.* 651, 682 (1979).

43. *Id.* at 683.

44. Mehrens & Phillips, *Sensitivity of Item Difficulties to Curricular Validity*, 24(4) *J. OF EDUC. MEASUREMENT* 357 (1987). See also Freeman, Belli, Porter, Floden, Schmidt, & Schwillie, *The Influence of Different Styles of Textbook Use on Instructional Validity of Standardized Tests*, 20(2) *J. OF EDUC. MEASUREMENT* 259 (1983); Airasian & Madaus, *Linking Testing and Instruction: Policy Issues*, 20(2) *J. OF EDUC. MEASUREMENT* 103 (1983); Leinhardt, *Overlap: Testing Whether It Is Taught*, in *THE COURTS, VALIDITY, AND MINIMUM COMPETENCY TESTING* 153 (G. Madaus ed. 1983).

45. Leinhardt, *supra* note 44, at 159.

46. McClung, *supra* note 42, at 682-83.

47. Popham & Lindheim, *Implications of a Landmark Ruling in Florida's Minimum Competency Test*, 63 *PHI DELTA KAPPAN* 18 (1981).

48. Linn, *supra* note 41, at 119 (emphasis in original). But see Pullin, *Minimum Competency Testing and the Demand for Accountability*, 63 *PHI DELTA KAPPAN* 20, 21 (1981) ("[T]he courts have decided that proof of fairness of a test requires evidence that *all* students have had a fair opportunity to learn the skills and knowledge covered by the test.").

pass the test.<sup>49</sup> However, if test exercises are used in instruction, the usefulness of the test as an instrument for measuring the student achievement is destroyed.<sup>50</sup> If there is too close a match between the instructional materials and the test, "the capacity to measure such important constructs as the understanding of a topic may be lost."<sup>51</sup> In addition, if the purpose of instruction is to teach students skills to apply to new problems, classroom instruction should not include problems identical to those on the test.<sup>52</sup>

Two major concerns in judging curricular or instructional validity are the closeness of the match required and whether materials from the previous grade should be included.<sup>53</sup> Overlap is measured by matching the test and curricula to a common scheme describing the substantive content. Thus, the degree of overlap between the tests and the curriculum is affected by the number of categories in the content classification scheme.<sup>54</sup> Further, because education is cumulative, the match will be greater across grades than it will be for a single grade.<sup>55</sup>

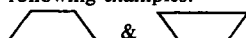


For example, one research team used a content classification scheme with 1260 categories to compare fourth grade mathematics textbooks with appropriate levels of major published standardized tests. The research

49. Freeman, Belli, Porter, Floden, Schmidt, & Schulle, *The Influence of Different Styles of Textbook Use on Instructional Validity of Standardized Tests*, 20(2) J. OF EDUC. MEASUREMENT 259 (1983).

50. Dyer, *Recycling the Problems in Testing. Proceeding of the 1972 Invitational Conference on Testing Problems*. Princeton, N.J.: Educational Testing Service (1973), quoted in Linn, *supra* note 41, at 126.

51. Linn, *supra* note 41, at 127.

52. *Id.* For example, suppose the concept to be learned is identification of a trapezoid. Suppose further that the students are given a definition, (a quadrilateral with one pair of parallel sides), and the following examples:

 &  . The student is then asked if this figure is a trapezoid:  If the student has learned the concept, the student should be able to apply the definition, look for one pair of parallel sides and determine that the figure is indeed a trapezoid. If the student was shown the figure in class and told it was a trapezoid, a correct answer might mean only that the student recalled what the teacher said, not that the student was applying the definition. See generally, Phillips & Mehrens, *Relating Achievement Test Scores and Item Statistics to Instructional Validity*, Paper presented at the National Council on Measurement in Education Annual Meeting, Washington, D.C. (April 1987).

53. Mehrens & Phillips, *Sensitivity of Item Difficulties to Curricular Validity*, 24(4) J. OF EDUC. MEASUREMENT 357 (1987).

54. *Id.* Mehrens & Phillips, *Detecting Impacts of Curricular Differences in Achievement Test Data*, 23(3) J. OF EDUC. MEASUREMENT 185, 186 (1986). The scheme for classifying content is often described in two dimensions: one dimension for types of numbers (such as whole numbers, fractions, and decimals) and the other dimension for numerical operations (such as addition or multiplication). The combination of a type of number and an operation (e.g., addition of fractions) identifies one cell in the two-dimensional classification table (also called a classification matrix). *Id.*

55. Mehrens & Phillips, *Sensitivity of Item Difficulties to Curricular Validity*, 24(4) J. OF EDUC. MEASUREMENT 357 (1987).



team found that the tests and textbooks had approximately 53%-74% common content.<sup>56</sup> However, when another research team made the same comparisons based on 180 categories (e.g., combining addition without carrying, addition with carrying, and column addition into a single category of addition), and combining textbook content for two adjacent grades, the percentage of common content between the tests and the textbooks increased to 81%-97%.<sup>57</sup>

If the curriculum test match is too close, testing for application and problem-solving skills is precluded.<sup>58</sup> If the match is *not* close enough, the courts will *not* be convinced that the test is fair.<sup>59</sup> Dr. Robert Linn summarized the dilemma: "[T]he challenge [is] to convince the courts that knowledge was taught — without precluding the possibility of measuring it."<sup>60</sup>

#### IV. *Debra P.* on Remand

The *Debra P.* remand hearing was held in early 1983, before the injunction prohibiting the use of the SSAT II to withhold diplomas expired.<sup>61</sup> On remand, the District Court determined: (1) whether the SSAT II "cover[ed] material was actually taught in Florida's classrooms"<sup>62</sup> and, (2) whether twelfth-grade black students' test scores were still being affected by vestiges of prior segregation in the Florida public schools.<sup>63</sup>

The Florida State Department of Education proffered a study by a private educational consulting firm to collect instructional validity data.<sup>64</sup> The firm collected and analyzed four kinds of data: (1) teacher survey responses regarding instruction on tested skills; (2) district surveys of skill placement, instructional materials, and test preparation activities; (3) site visit reports designed to verify the district reports; and (4) student surveys

---

56. Schmidt, *Validity as a Variable: Can the Same Certification Test Be Valid for All Students?*, in *THE COURTS, VALIDITY, & MINIMUM COMPETENCY TESTING* 133, 144 (G. Madaus ed. 1983).

57. Mehrens & Phillips, *supra* note 44, at 364-65.

58. Linn, *supra* note 41, at 128.

59. *Id.*

60. *Id.* at 129.

61. *Debra P. v. Turlington*, 564 F. Supp. 177, 179 (M.D. Fla. 1983).

62. *Id.* In an earlier decision, a Georgia court implemented the *Debra P.* mandate for content validity when it accepted expert and teacher testimony to show that the skills tested were in the curriculum and being taught in the schools in the district. *Anderson v. Banks*, 520 F. Supp. 472 (S.D. Ga. 1981), *reh'g* 540 F. Supp. 761 (S.D. Ga. 1982).

63. *Debra P.*, 564 F. Supp. at 179. In 1983, graduating seniors had had six years' notice of the testing requirement, six years of instruction targeted at the skills tested and had never attended segregated schools. *Debra P. v. Turlington*, 730 F.2d 1405, 1407 (11th Cir. 1984).

64. *Debra P.*, 564 F. Supp. at 180.

asking sampled groups of students whether they had been taught the specific skills tested.<sup>65</sup>

The study indicated that each student received mastery level instruction on the tested skills an average of 2.7 times, which exceeded the single mastery level exposure required for an instructionally valid and fair test.<sup>66</sup> The plaintiffs were concerned that instructions had *not* been documented for all 12 years of the students' education (that the study was cross-sectional rather than longitudinal), that the survey responses did *not* accurately reflect classroom activities for all students, and that the survey results were inadmissible hearsay.<sup>67</sup>

The district court held that the state had met its burden of demonstrating instructional validity, that the SSAT II was a fair test of what was taught, and that the SSAT II could be used prospectively as a diploma sanction test.<sup>68</sup> The court further held that, even if the present disproportionate failure rates were caused by past discrimination, the state had adequately demonstrated that the SSAT II was a necessary remedy.<sup>69</sup> In so holding, the court noted that it was not constitutionally unfair that some students had mediocre teachers.<sup>70</sup> The court also noted that proving instructional validity for each individual student was an impossible burden.<sup>71</sup> The appropriate standard for instructional validity was that "the skills be included in the official curriculum and that the majority of the teachers recognize them as being something they should teach."<sup>72</sup> The district court decision was affirmed on appeal.<sup>73</sup>

## V. Criticism of Diploma Sanction Tests

Proponents make eight arguments in favor of diploma sanction tests while opponents advance eleven arguments against the tests.<sup>74</sup> Pro-

---

65. *Id.* at 180-81. For additional details of the study, see Fisher, *Implementing an Instructional Validity Study of the Florida High School Graduation Test*, 2(4) EDUC. MEASUREMENT: ISSUES & PRAC. 8 (1983).

66. *Id.* at 180.

67. *Id.* at 182.

68. *Id.* at 186. The study results were admitted under the residual hearsay exception as the basis for the experts' opinions. *Id.* at 182-83, n. 7.

69. *Id.* at 188.

70. *Debra P. v. Turlington*, 564 F. Supp. 177, at 183-85. The *Anderson* court previously noted that "further debate about the efficacy of the program belongs in the School Board meetings and not in the Court." *Anderson v. Banks*, 540 F. Supp. 472 (S.D. Ga. 1981), *reh'g* 540 F. Supp. 761, 766 (S.D. Ga. 1982).

71. *Debra P.*, 564 F. Supp. at 185-86 (M.D. Fla. 1983).

72. *Id.* at 186.

73. *Debra P. v. Turlington*, 730 F.2d 1405 (11th Cir.) (1984).

74. Note, *supra* note 8, at 1116-18.

ponents' arguments include: (1) provision of academic meaning of diplomas beyond prescribed "seat time;" (2) setting of explicit standards in place of social promotion; (3) return to basics to counter innovative curricular fluff; (4) diagnosis for student remediation; (5) establishment of a proficiency requirement for minority students who might otherwise pass through the system without adequate learning; (6) motivation of students to learn; (7) provision of a standard for public accountability; and (8) improvement of the public image of education among the citizenry to enhance the chances for approval of increased financial support for education.<sup>75</sup>

Counterarguments focus on the fact that the tests solve fewer problems than they create.<sup>76</sup> The counterarguments include: (1) the basing of an important decision on a single variable; (2) too little phase-in time for schools to teach the tested skills; (3) political rather than academic determination of passing standards; (4) restriction of the curriculum to teaching only the tested skills; (5) the conversion of minimums to maximums to the detriment of educational excellence; (6) the inadequacy of written tests to measure life skills; (7) the unjustifiability of high costs of test development and remediation; (8) increased dropout rates; (9) the students' having to bear the stigma for the schools' failure to educate; (10) the disproportionate share of the burden of failure falling on minority students; and (11) the fact that tests are a simplistic solution to the historical problem of underachievement.<sup>77</sup> Arguments (2), (3), and (7) can be addressed through program changes, delayed implementation, and increased efficiency; but some of the others require a more detailed analysis.<sup>78</sup>

### A. *Blaming the Tests*

In his 1986 Presidential Address to the National Council on Measurement in Education, Dr. William Mehrens observed:

[M]easurement critics do not always differentiate between the problems that tests cause and the problems that tests reveal. We have, in our society, an overwhelming tendency to kill the messenger who brings the bad news. . . . [I]n the minimum competency testing that takes place in the public schools in this nation, it has been found that some high school seniors cannot do eighth-grade work. [Testing] advocates say this indicates inadequate learning. . . . [C]ritics say this indicates invalid testing. To paraphrase a line from Jeane Kirkpatrick's speech at

---

75. *Id.*

76. *Id.*

77. *Id.*

78. See generally, *Debra P. v. Turlington*, 730 F.2d 1405 (11th Cir. 1984); Citron, *Courts Provide Insight on Content Validity Requirements*, 2(4) EDUC. MEASUREMENT: ISSUES & PRAC. 6 (1983) ("inadequate notice is a curable deficiency").

the 1984 Republican convention: But then, somehow, . . . critics *always* blame the tests first.<sup>79</sup>

The policy decisions educators make using test data will be made whether or not the test data is available.<sup>80</sup> Someone will decide whether students have met district criteria in order to receive a high school diploma.<sup>81</sup> Thus, it is not the test that is high stakes, but the decisions being made. Those decisions will continue to be made regardless of whether a test is given or *not*.<sup>82</sup> The more data one uses to make a decision, the more accurate the decision.<sup>83</sup> If test developers have constructed a reliable and valid test the test will contribute to the accuracy of decisions made, even when testing experts can suggest ways of improving a test given additional resources.<sup>84</sup>

### ***B. Conjunctive vs. Compensatory Models***

With respect to the criticism that the decision is being made on a single piece of data when diploma sanction tests are given, Dr. Mehrens responded: "[The issue is not] whether more data lead to better decisions than fewer data. They do. . . . [The issue] is how we should combine data."<sup>85</sup>

There are two models that are commonly used to combine data: conjunctive and compensatory.<sup>86</sup> The conjunctive model requires adequate performance on each measure, and the data is typically collected sequentially to maintain cost effectiveness.<sup>87</sup> Because failure to meet the standard on any one measure eliminates that individual from further consideration, there is *no* point in incurring the expense of collecting additional irrelevant data.<sup>88</sup> However, when the compensatory model is used, *all* data must be collected on *all* individuals because performance above the standard on one measure can compensate for substandard performance on another.<sup>89</sup>

---

79. Mehrens, *Measurement Specialists: Motive to Achieve or Motive to Avoid Failure?*, 5(4) EDUC. MEASUREMENT: ISSUES & PRAC. 5, 7 (1986) (emphasis in original). See also Fisher, *A Reaction to Mehrens's NCME Presidential Address*, 5(4) EDUC. MEASUREMENT: ISSUES & PRAC. 15 (1986) (supporting Mehrens' position).

80. Mehrens, *Keeping the Faith: A Response to the Reactors*, 5(4) EDUC. MEASUREMENT: ISSUES & PRAC. 17 (1986).

81. *Id.*

82. *Id.*

83. W. MEHRENS & I. LEHMANN, *USING STANDARDIZED TESTS IN EDUCATION* 10 (4th ed. 1987).

84. Mehrens, *Measurement Specialists: Motive to Achieve or Motive to Avoid Failure?*, 5(4) EDUC. MEASUREMENT: ISSUES & PRAC. 5 (1986).

85. *Id.* at 7-8.

86. *Id.*

87. *Id.*

88. *Id.*

89. *Id.*

Multiple sources of data are imperative to the decision of whether or not a particular student should be awarded a high school diploma.<sup>90</sup> Test scores are *not* the only data collected.<sup>91</sup> Students must also pass all required courses.<sup>92</sup> According to the conjunctive model, a student who passes the test will *not* receive a diploma until all the required coursework is completed. Further, a student who completes all the required coursework will *not* receive a diploma until the test is passed.<sup>93</sup> Both pieces of data, the test and the required course work, are used to make the decision as to whether a student graduates.<sup>94</sup>

Critics argue that the model chosen should be compensatory rather than conjunctive.<sup>95</sup> According to the compensatory model, students who receive low test scores receive a diploma if they pass all of their required high school courses.<sup>96</sup> But the compensatory model is appropriate only if it is reasonable to assume that passing typical high school classes such as physical education, typing, home economics, or history compensates for a lack of basic skills in mathematics and communication.<sup>97</sup> If one believes that students must be able to figure sales tax or make change in order to function adequately in the everyday world, anything less than mastery of these skills is unacceptable.<sup>98</sup> Alternatively, if it is argued that the coursework will remediate deficiencies, the student should pass the test on a subsequent administration.<sup>99</sup>

---

90. *Id.* See also Popham, *Task-Teaching Versus Test-Teaching*, 2(4) EDUC. MEASUREMENT: ISSUES & PRAC. 10, 11 (1983) (analogizing denial of a diploma due to failure of a statewide test to denial of a diploma based on failure to pass a required course due to the student's failing a final exam).

91. See, e.g., *Debra P. v. Turlington*, 474 F. Supp. 244, 247 (M.D. Fla. 1979).

92. *Id.*

93. See Mehrens, *supra* note 84, at 8.

94. *Id.*

95. *Id.*

96. See generally Mehrens, *supra* note 84, at 8.

97. *Id.*

98. *Id.*

99. See *Debra P. v. Turlington*, 564 F. Supp. 177, 185 (M.D. Fla. 1983), *aff'd*, 730 F.2d 1405, 1410-11 (11th Cir. 1984). Note that some states set their passing standard for teacher competency tests as much as "three standard errors of measurement *below* the minimum standards determined from their standard-setting studies. . . . Given that a person has repeated opportunities to take the test, there is virtually *no* chance that a person whose true score is above the judged minimum standard would not [pass]. However, after three attempts, 87.5% of those whose true score is three standard errors below the judged minimum standard would pass the test." Mehrens, *supra* note 84 at 9 (emphasis in original). Thus, with repeated attempts to take a diploma sanction test following remediation, a student with a true ability above the minimum standard would almost certainly pass and even some students with true ability just below the passing score on the functional literacy test (i.e., the students had not mastered the skills tested) would have a substantial probability of passing. *Id.*

## VI. New Problems From Old Solutions

The *Debra P.* case has substantially affected the implementation of graduation tests in other states.<sup>100</sup> Due to the institution of testing requirements many marginal students have received the basic education they had not previously received.<sup>101</sup> Basic skills became an instructional priority after state education departments published lists of specific objectives to be tested. Some states made old test forms available to teachers.<sup>102</sup> This satisfied the due process notice mandated by the courts, but had some unexpected side effects.<sup>103</sup> In some schools, teachers spent the entire school year drilling students exclusively on the 10, 12 or 15 objectives covered by the test.<sup>104</sup> Further, some teachers used practice items identical or nearly identical to the items on the test.<sup>105</sup> As a result of preparing students for the test, other portions of the mandated state curriculum were often neglected.<sup>106</sup>

### A. National Comparisons

At the time criticisms of excessive and inappropriate test preparation emerged, state officials became concerned about how the achievement of students in their state compared to national achievement.<sup>107</sup> State officials

---

100. See Calfee, *Standards, Evidence and Equity: Implications of the 1983 Debra P. Decision*, 2(4) EDUC. MEASUREMENT: ISSUES & PRAC. 11, 12 (1983) ("Other states now have a precedent and can proceed with the implementation of similar programs with assurance that they are unlikely to be challenged in court."); Lott, *A Helpful Decision in an Important Case*, 2(4) EDUC. MEASUREMENT: ISSUES & PRAC. 13 (1983) ("The decision is also good news for the testing specialists in the states where competency testing litigation is pending or likely; they now have a precedent for demonstrating that their competency tests have curricular and instructional validity.").

101. See Popham, *Task-Teaching Versus Test-Teaching*, 2(4) EDUC. MEASUREMENT: ISSUES & PRAC. 10, 11 (1983).

102. Rothman, *Texas Officials Modify State Assessment to Minimize Ways to "Teach to the Test"*, EDUC. WK., March 29, 1989, at 7 col. 1; Interview with Stanley Rabinowitz, Director of Student Assessment, New Jersey State Department of Education (July 5, 1989). See also, Note, *supra* note 10, at 878 ("What the [*Debra P.*] court was actually saying was that students must be told what parts of their education are relevant [to the test] (and therefore to be remembered), and what parts are in effect useless [for passing the test] (and may therefore be forgotten as soon as they are taught).").

103. Texas Education Agency, *A Customized Test for Texas 2* (June, 1989) (draft).

104. See Popham, *supra* note 101 at 11 (warning that overemphasis on teaching to the test might create "curriculum reductionism"); Rothman, *Texas Officials Modify State Assessment to Minimize Ways to "Teach to the Test"*, EDUC. WK., March 29, 1989, at 7, col. 1, 2.

105. Interview with Keith Cruse, Director of Student Assessment, Texas Education Agency (June 29, 1989); Texas Education Agency, *A Customized Test for Texas 2* (June 1989) (draft).

106. Rothman, *Texas Officials Modify State Assessment to Minimize Ways to "Teach to the Test"*, EDUC. WK., March 29, 1989, at 7, col. 1, 2.

107. Interview with Keith Cruse, Director of Student Assessment, Texas Education Agency, June 29, 1989; Texas Education Agency, *A Customized Test for Texas 4* (June 1989) (draft).

noted that, even though in successive years their students improved on the state objectives tested, they might still be achieving at a lower level than students in other states.<sup>108</sup> This concern, together with legislative mandates for normative scores, created a need for national norms for the state test.<sup>109</sup> One method for obtaining national norms for the statewide test is to equate the state test to a nationally standardized achievement test. When this method is used, the national norms are most appropriately interpreted as an estimate of how a national sample of students would have performed on the statewide objectives.<sup>110</sup>

Many educators misinterpreted the comparison, believing that the normative information was an estimate of how students in the state would have performed on the nationally-standardized achievement test.<sup>111</sup> Since educators in the state were concentrating on teaching the narrow range of skills included on the the statewide test (and *not* the broader set of skills tested on the standardized achievement test), increases in scores on the statewide test resulted in what appeared to be concomitant increases in national standing.<sup>112</sup>

The national gains were largely spurious.<sup>113</sup> Followup studies showed that actual increases in achievement test performance were much smaller than predicted by the increased performance on the statewide test.<sup>114</sup> However, state officials did not want to administer a new standardized achievement test to all students in the state.<sup>115</sup> Thus, because of concerns that there was already too much testing, they continued to search for ways to obtain normative data while keeping the central focus of the statewide testing program on the statewide test.<sup>116</sup>

### B. The Lake Wobegon Effect

About the same time state officials were debating about how best to obtain national norms for their statewide tests, Dr. John Cannell discovered

---

108. *Id.*

109. *Id.* See generally, Yen, Green & Burket, *Valid Normative Information from Customized Achievement Tests*, 6(1) EDUC. MEASUREMENT: ISSUES & PRAC. 7 (1987).

110. *Id.*

111. Interview with Keith Cruse, Director of Student Assessment, Texas Education Agency (June 29, 1989).

112. Yen, Green & Burket, *Valid Normative Information from Customized Achievement Tests*, 6(1) EDUC. MEASUREMENT: ISSUES & PRAC. 7, 10 (1987).

113. *Id.*; Texas Education Agency, *A Customized Test for Texas 6* (June 1989) (draft).

114. Yen, Green & Burket, *Valid Normative Information from Customized Achievement Tests*, 6(1) EDUC. MEASUREMENT: ISSUES & PRAC. 7, 10 (1987).

115. Interview with Keith Cruse, Director of Student Assessment, Texas Education Agency (June 29, 1989).

116. *Id.*; see generally, Yen, Green & Burket, *Valid Normative Information from Customized Achievement Tests*, 6(1) EDUC. MEASUREMENT: ISSUES & PRAC. 7 (1987).

the "Lake Wobegon effect" in public education.<sup>117</sup> In Garrison Keillor's long-running radio show, *A Prairie Home Companion*, Mr. Keillor popularized a mythical town in Minnesota, Lake Wobegon, where all students "are above average."<sup>118</sup> Recent data suggests that the nation is making progress toward a Lake Wobegon utopia.<sup>119</sup>

Dr. Cannell surveyed states with mandated nationally-normed standardized achievement tests and the large urban districts in states where standardized tests are chosen at the district level.<sup>120</sup> Nearly all reported that their students were achieving above the national average.<sup>121</sup> Dr. Cannell's home state of West Virginia ranked low on every other educational indicator but its sixth grade students were at the sixty-fifth percentile on the standardized test.<sup>122</sup> Assuming these data were correct, the average West Virginia sixth grader scored higher than 65% of all students in the national norm group.<sup>123</sup> In addition, some educators inferred from the data that the average West Virginia sixth grade student's achievement was higher than 65% of all sixth graders in the nation.<sup>124</sup> By contrast, Michigan, which scores near the national average on per capita income, college entrance scores, poverty rate, graduation rate, etc.,<sup>125</sup> reported above average performance for 88% of its twenty largest districts plus twenty others selected at random.<sup>126</sup>

Dr. Cannell suggested two possible explanations for the Lake Wobegon phenomena: inaccurate norms and teaching the test.<sup>127</sup> These two ex-

---

117. Cannell, *Nationally Normed Elementary Achievement Testing in America's Public Schools: How All 50 States Are Above the National Average*, 7(2) EDUC. MEASUREMENT: ISSUES & PRAC. 5 (1988). Dr. Cannell is the founder and President of Friends For Education, the group which sponsored the research.

118. Haney, *Editorial*, 7(4) EDUC. MEASUREMENT: ISSUES & PRAC. 4 (1988).

119. Mehrens & Kaminski, *supra* note 13.

120. Cannell, *supra* note 117, at 5-6.

121. *Id.* at 6-7. A national panel of educators, appointed by the U.S. Department of Education, recently confirmed Dr. Cannell's results.

122. Cannell, *supra* note 117.

123. P. BLOMMERS & R. FORSYTH, *ELEMENTARY STATISTICAL METHODS IN PSYCHOLOGY AND EDUCATION* 51 (2nd ed. 1977); W. MEHRENS & I. LEHMANN, *USING STANDARDIZED TESTS IN EDUCATION* 113-14 (4th ed. 1987).

124. P. BLOMMERS & R. FORSYTH, *ELEMENTARY STATISTICAL METHODS IN PSYCHOLOGY AND EDUCATION* 51 (2nd ed. 1977). However, such an inference would be accurate only if the national norm group were representative of current student performance nationally. As Dr. Cannell put it, "An above-average score does not mean that the student or the district or the state is above the current year's national average. It means only that the score achieved is better than the mean score achieved by the norm group in years past." Cannell, *supra* note 117 at 7.

125. Cannell, *supra* note 117 at 7 (using data from the U.S. Dep't of Educ., Office of Planning, Plan, Budget, and Evaluation, *State Educ. Statistics Supplement: Student Performance and Resource Inputs*, 1985-1986 (1987)).

126. *Id.*

127. *Id.* The U.S. Department of Education provided several hypotheses for the phenomenon:



planations are related. While the normative data may have been accurate when collected, it is no longer appropriate, when standardized testing conditions have been violated by directly teaching the test.<sup>128</sup>

### C. Teaching the Test

Dr. Cannell stated that: "Questions on these [standardized achievement] tests change only every 8 to 10 years; thus, [teachers] can easily teach the test because they administered it the year before."<sup>129</sup> Furthermore, in states where steps have been taken to prevent the teaching of the tests, dramatic yearly gains are not found."<sup>130</sup> Examples of survey data collected from Utah, Wisconsin, California and Maine, where strict measures insure that the test is not taught, revealed only small annual increases in test scores.<sup>131</sup> Dr. Cannell concluded:

If the test is given under strict security, not used for curriculum purposes, and not teachable, then the only method that below-average states will have to improve scores will be to improve curriculum, attract better teachers with higher salaries, improve remediation programs, fire incompetent teachers and principals, buy

---

(1) unrepresentative national norms; (2) choosing the test most closely aligned (matched) with local curricula; (3) the tendency of the students taking the test to be more motivated than those in the norming sample were; (4) the too infrequent updating of norms; (5) use of identical test questions for several years in the same school system; (6) noncomparability of user samples to norm samples. Phillips & Finn, *The Lake Wobegon Effect: A Skeleton in the Testing Closet?*, 7(2) EDUC. MEASUREMENT: ISSUES & PRAC. 10, 10-11 (1988). It is interesting that the Department did not mention the phenomenon of teaching to the test documented by Cannell, *supra* note 117 at 8-9 and Mehrens & Kaminski, *supra* note 13. To those who claim the test scores represent real increases in achievement, Dr. Cannell cites high school data from a variety of sources that contradict the large gains reported at the elementary level. He also references 1985 NAEP data indicating that 80% of the blacks educated in the last 15 years are functionally illiterate — they cannot read a map or a bus schedule or balance a checkbook. Cannell, *The Lake Wobegon Effect Revisited*, 7(4) EDUC. MEASUREMENT: ISSUES & PRAC. 12, 14 (1988). See also Koretz, *Arriving in Lake Wobegon: Are Standardized Achievement Tests Exaggerating Advancement and Distorting Instruction?*, 12(2) AM. EDUCATOR 8, 10 (1988) (suggesting that educators and policy makers have been misled into believing that only the secondary schools are at risk).

128. W. MEHRENS & I. LEHMANN, USING STANDARDIZED TESTS IN EDUCATION 148-49 (4th ed. 1987). One test publisher is offering national norms updated annually based on user samples. While the sample may not be strictly representative nationally, approximately half the students should score below average. Cannell, *supra* note 117, at 9. Although Dr. Cannell sees this procedure as a partial solution, educators must recognize that user samples have even more problems with representativeness than those described by the U.S. Department of Education for regular norming samples. Phillips & Finn, *The Lake Wobegon Effect: A Skeleton in the Testing Closet?*, 7(2) EDUC. MEASUREMENT, ISSUES & PRAC. 10, 10-11 (1988). Yearly norms also have the disadvantage that schools using them are unable to measure longitudinal changes. Stonehill, *Norm-Referenced Test Gains May Be Real: A Response to John Jacob Cannell*, 7(2) EDUC. MEASUREMENT: ISSUES & PRAC. 23, 24 (1988).

129. Cannell, *supra* note 117, at 8.

130. *Id.* at 8-9.

131. *Id.*

textbooks that are more difficult, increase graduation requirements, and spend more money on education.<sup>132</sup>

Mehrens & Kaminski also contend that teaching the test instead of the skills is a significant problem. "We *strongly* suspect a major reason for the high scores is teaching the test — this practice has increased because of the increasing efforts to use scores as measures of school and teacher quality."<sup>133</sup> In opposition, Wrightstone asserts that it is unethical to use any parallel forms of the test in instruction while Cohen claims that the instruction and the test items should be identical so that they will result in "near perfect learning."<sup>134</sup> The fallacy with Cohen's approach is that teaching the test makes it impossible to generalize from the test items to similar tasks in order to gauge the student's level of understanding and ability to apply what was learned.<sup>135</sup> Mehrens and Kaminski analyzed several types of commercial test preparation materials for standardized achievement tests that cover the range between teaching the skills but *not* the test items and teaching similar or exactly the same test items.<sup>136</sup> They found that many of the materials matched the test items too closely, thus discrediting educators' inferences that high test scores indicated that students had learned the concepts tested.<sup>137</sup>

Mehrens and Kaminski noted that, in one district, 11% of the teachers did not believe that teaching the test items was "cheating."<sup>138</sup> Dr. Robert Linn observed: "For some of those who are philosophically opposed to a [basic skills test] requirement for graduation, cheating may be seen as a case of the ends justifying the means."<sup>139</sup>

## VII. New Solutions and Legal Challenges

The identification of the Lake Wobegon effect is important in the

---

132. *Id.* at 9.

133. Mehrens & Kaminski, *supra* note 13 at 14 (emphasis in original). *See also*, Popham, *supra* note 90, at 10-11 (if task descriptions clear, teachers can teach task rather than test items).

134. Mehrens & Kaminski, *supra* note 13, at 15. *See also* Cohen, *Instructional Alignment: Searching for a Magic Bullet*, 15(8) EDUC. RESEARCHER 16, 16-20 (1987).

135. Mehrens & Kaminski, *supra* note 13 at 15.

136. *Id.* at 16-18.

137. *Id.* at 21. For example, one test preparation program told potential examinees that the four options on a multiple-choice question on distinguishing fact from opinion would contain 2 fact and 2 opinion choices. Further, they were told that one of each of the types would be in the passage and one would not. Dr. Shepard characterized this as "exploitation of the test specifications." Shepard, *A Case Study of the Texas Teacher Test: Technical Report* (1987).

138. Mehrens & Kaminski, *supra* note 13, at 18.

139. Linn, *supra* note 41, at 127. *See also*, Popham, *supra* note 90 ("[F]rom reading the decisions of the *Debra P. v. Turlington* trial, educators might reasonably conclude that not only is teaching to the test now acceptable, it is downright mandatory.").

diploma testing arena for two reasons: (1) it raises concerns about the accuracy of normative comparisons based on linking the statewide test to standardized achievement tests and (2) if teachers invalidate a standardized test by teaching the test questions, they may also invalidate the diploma test.<sup>140</sup> In reacting to the Lake Wobegon effect, states have once again revised their testing programs.<sup>141</sup>

### *A. Sampling Objectives*

Proposed revisions include sampling the statewide objectives so that the skills tested from year to year will not be the same.<sup>142</sup> The aim is to force teachers to teach all of the objectives in the statewide curriculum rather than just the few that are tested.<sup>143</sup> Further, state norms may be more accurately linked to national norms if questions from the for-sale versions of the nationally-normed achievement tests which were developed at the same time, and have never been released for public use, are used in state testing.<sup>144</sup>

### *B. Legal Challenges to New Solutions*

Proposed solutions are scheduled for implementation in the early 1990s.<sup>145</sup> Once implemented, the proportion of students who fail will initially increase and a disproportionate number of failures may occur among minority students.<sup>146</sup> Once again, courts will be asked if the testing procedures are "instructionally valid" or if they violate the equal protection rights of minority students.<sup>147</sup>

## **VIII. Should the Courts Find the New Solutions Constitutional?**

The court in *Goss* recognized a property right to a public education but not specifically to a high school diploma.<sup>148</sup> However, past notes, com-

---

140. See generally, Cannell, *supra* note 117; Mehrens & Kaminski, *supra* note 13; Rothman, *supra* note 106.

141. Rothman, *supra* note 106.

142. *Id.* at col. 2.

143. *Id.*

144. Texas Education Agency, *A Customized Test for Texas* 10 (June 1989) (draft).

145. Rothman, *supra* note 106.

146. See Pryor, *Student Competency Testing in Texas*, 16 ST. MARY'S L. J. 903, 921. Since the average scores of minorities tend to be lower, when the tests become harder or the standard for passing is increased, a greater number of minority students will be affected. *Debra P. v. Turlington*, 474 F. Supp. 244, 249 (1979).

147. See generally the *Debra P.* cases, *supra* note 16.

148. *Goss v. Lopez*, 419 U.S. 565, 574 (1975). See also, Note, *supra* note 10, at 879-80, n. 70.

ments, and articles view minorities' failure to receive a diploma due to failing the test as deprivation of their property rights. Further, their liberty rights are also deemed violated because, in failing the tests, minority students are stigmatized for life.<sup>149</sup> Critics also viewed the test as the only hurdle standing in the way of graduation, despite other state and local district requirements.<sup>150</sup>

### A. *What Is Best For Minorities?*

When the equal protection question is revisited, courts should consider that<sup>151</sup> many minority students pass the graduation tests and employers consequently know that those students have mastered certain skills.<sup>152</sup> Even if the test were not given, those incapable of passing it will have difficulties performing their jobs; and employers will expend more time and resources identifying these skill deficiencies.<sup>153</sup> And, in the opinion of some prominent black educators, minority students are insulted by suggestions that they cannot learn the skills tested.<sup>154</sup>

### B. *Equal Opportunity Versus Equal Outcome*

Many confuse the constitutional right of equal education<sup>155</sup> with the nonexistent right of equal outcome.<sup>156</sup> The constitution does not guarantee that all persons will achieve at the same level as a result of their education.<sup>157</sup> "It has only been in the past 20 years or so that business en-

149. See, e.g., McClung, *supra* note 42; Note, *Functional Literacy Testing and the Denial of High School Diplomas in a Post-Desegregation Setting: Debra P. v. Turlington*, 33 RUTGERS L. REV. 564 (1981); Note, *supra* note 8; Note, *Education and the Law: Functional Literacy Program — A Matter of Timing*, 10 STETSON L. REV. 125 (1980); GUNN, *Debra P. v. Turlington: Due Process Enters the Classroom, but How Far?*, 11 J. OF L. & EDUC. 573 (1982).

150. McCLUNG, *supra* note 42.

151. See generally, Lerner, *The Minimum Competence Testing Movement*, 36(10) AM. PSYCHOLOGIST 1057 (1981).

152. See *supra* note 80; *Debra P. v. Turlington*, 474 F. Supp. 244, 248-49 (M.D. Fla. 1979).

153. Note, *Constitutional Law — Discrimination — Passage of Functional Literacy Examination as Prerequisite to Receipt of High School Diploma Held Unconstitutional*, 10 CUMB. L. REV. 863, 879, n. 69 (1980).

154. Raspberry, *Top Black Scholars Proof of Excellence*, Lansing State J., May 20, 1988, at 4, col. 2 (citing a study that found black student peer pressure *not to study because it was equated with "acting white."*).

155. *Brown v. Board of Education*, 347 U.S. 483 (1954).

156. *Debra P. v. Turlington*, 654 F.2d 1079, 1082 (5th Cir. 1981) (Tjoflat, J., dissenting). See also, Note, *supra* note 10, at 879.

157. *Debra P.*, 654 F.2d at 1082. (Tjoflat, J., dissenting). But see, Note, *Functional Literacy Testing and the Denial of High School Diplomas in a Post-Desegregation Setting: Debra P. v. Turlington*, 33 RUTGERS L. REV. 564, 589 (1981) (arguing that the injunction in the *Debra P.* case should have continued until blacks no longer had difficulty achieving academically or until they achieved peer status, which Frizzell defined as when blacks achievement equals that of whites).

trepreneurs have capitalized on the idea of marketing tutoring and other educational services on a grand scale. . . . Many observers worry, however, about the fairness of allowing some students to get such extra help while others' parents cannot afford it no matter how strong their commitment to education. . . . But others say the disparities may be unavoidable."<sup>158</sup> According to Robert R. Spillane, Superintendent of Schools in Fairfax County, Virginia, "That is a reality in a democratic, competitive society. . . . What we should do is balance up rather than level down."<sup>159</sup>

### C. *Dissenting Views From the District Court*

The appeals court denied a petition in *Debra P.* for rehearing, and Judge Tjoflat, dissenting, stated that the court "held it *constitutionally* impermissible for the State of Florida to presently require the same level of functional literacy from black and white high school students."<sup>160</sup> Judge Tjoflat further argued that *Goss*,<sup>161</sup> *which specified minimum due process protections for students*, should apply only to disciplinary actions interfering with a student's *access* to education and *not* diploma denial based on academic decisions.<sup>162</sup> While property rights require a "full and fair opportunity for an education," academic success in that education is not a right, but a reward to be earned.<sup>163</sup> As Judge Hill, in his dissent to *Debra P.*, expressed, "the child's expectation that he or she will receive a diploma without having achieved even a minimum degree of education is

---

158. Walsh, *Private Education Firms Discovering A Lucrative Market in Public Schools: Schools "Could Do Just What We Do If They Had the Same Resources,"* EDUC. WK., June 7, 1989, at 7.

159. *Id.*

160. *Debra P. v. Turlington*, 654 F.2d 1079, 1081 (5th Cir. 1981) (Tjoflat, J., dissenting) (emphasis in original).

161. *Goss v. Lopez*, 419 U.S. 565 (1975).

162. *Debra P.*, 654 F.2d at 1081-82. See also *Board of Curators of University of Missouri v. Horowitz*, 435 U.S. 78 (1978) (holding that the *Goss* minimum due process requirements apply only to disciplinary decisions limiting educational access); *Gaspar v. Bruton*, 513 F.2d 843, 850 (10th Cir. 1975); *Mahavongsanan v. Hall*, 401 F. Supp. 381, 383 (N.D. Ga. 1975) *rev'd*, 529 F.2d 448 (5th Cir. 1976) (holding disciplinary but *not* academic decisions of educational institutions judicially reviewable). Note that Judge Hill also dissented in the decision denying rehearing. Judge Hill wrote a dissenting opinion that supported that of Judge Tjoflat and concurred in Judge Tjoflat's analysis of relevant case law. Throughout the following discussion, the reader is encouraged to also consult Judge Hill's dissent for further support. *Debra P.*, 654 F.2d at 1086-89.

163. *Debra P.*, 654 F.2d at 1082. See also, *supra* note 10, at 879 (suggesting that the property right the black students obtained under the *Debra P.* decision was the diploma itself and *not* the education to which they were entitled because possession of the diploma was *no* guarantee of mastery of the requisite basic skills).

held [by the majority] to be a constitutionally protected property right."<sup>164</sup>

Judge Tjoflat further asserted that success on a functional literacy exam involving common everyday tasks (e.g., price discounts, comparison-shopping) requires a combination of basic skills and maturation.<sup>165</sup> Requiring a state to demonstrate that all students were taught well and that all teachers dealt with maturation issues places an onerous burden on a state trying to improve the quality of education.<sup>166</sup> Judge Tjoflat predicted that the appeals court decision was an incentive to all states to return to a system of social promotion.<sup>167</sup>

Judge Tjoflat also stated that, contrary to *Brown*,<sup>168</sup> the majority used the fourteenth amendment to sanction unequal treatment of black and white students by insisting that illiterate blacks be given a diploma.<sup>169</sup> He asserted that such actions made the diploma a worthless piece of paper and perpetuated the widespread belief that blacks are incapable of achieving at the same level as white students.<sup>170</sup> Judge Tjoflat characterized the federal court intervention as a paternalistic attempt to correct past discrimination by giving away meaningless credentials rather than insisting on real educational achievement.<sup>171</sup> He further suggested that such ac-

---

164. *Debra P.*, 654 F. 2d at 1087.

165. *Id.* at 1082.

166. *Id.* at 1083.

167. *Id.* at 1084.

168. *Brown v. Board of Education*, 347 U.S. 483 (1954). Judge Hill observed, "Instead, we endorse the charge levelled against the courts and solemnly find that the Constitution and the law mandate that quality education, since *Brown*, be forbidden." *Debra P.*, 654 F.2d at 1089.

169. *Debra P. v. Turlington*, 654 F.2d 1079, 1084-85 (5th Cir. 1981). It is noteworthy in this context that there seemed to be little disagreement that the test measured basic functional skills and that the students who failed did *not* have those skills. Thus it appeared that by court order some students would be certified as having satisfactory educational attainment when they were in fact functionally illiterate. It also seemed likely that the benefit to the student would only be temporary; any employer relying on the diploma would certainly discover the lack of skills in time. Meanwhile, the student would have false hopes of a better job and ultimately face the disappointment and stigmatization of losing the job. By requiring Florida to award degrees that had not been earned, the court simply delayed but did not change the reality that those students were unprepared to enter the job market. *Id.* But see Note, *Education and the Law: Functional Literacy Program — A Matter of Timing*, 10 STETSON L. REV. 125, 136-37 (1980) (arguing that the majority decision was an extension of *Brown* that prevented punishment of the victims of past discrimination).

170. *Debra P.*, 654 F.2d at 1084-85. The majority did seem to be overlooking the fact that 80% of the black graduating seniors had *passed* the test and that by the third try, the proportion of black and white students failing the test was nearly identical. *Debra P. v. Turlington*, 474 F. Supp. 244, 248-49 (M.D. Fla. 1979). See also Judge Hill's dissent in *Debra P. v. Turlington*, 654 F.2d 1079, 1088 (5th Cir. 1981) ("It is undoubtedly true that the appearance of having been educated may be accomplished by the conferring of a diploma. Nevertheless, if the child has received no learning, even the most emphatic judgment and order of the most diligent court cannot supply it.").

171. *Debra P. v. Turlington*, 654 F.2d 1079, 1084-85 (5th Cir. 1981). Judge Hill characterized this as "postpon[ing] a quality education for three years. . . . [A] pupil's achievement in an educa-

tions encouraged blacks *not* to achieve because diplomas would be available to them no matter what their level of achievement.<sup>172</sup> Real equality, the Judge insisted, can be achieved only when those who have *not* met the standard are identified and retained until they pass the test and legitimately obtain their degrees.<sup>173</sup>

#### D. *The Student's Responsibility*

Recent Asian immigrants have succeeded remarkably in American schools.<sup>174</sup> Many refugees come to the U.S. with only the clothes on their backs, begin their education speaking little if any English, and finish school at the top of their class.<sup>175</sup> Except for very recent immigrants, few receive public assistance.<sup>176</sup> The Asian-American immigrants begin earning less than their white counterparts but, within a decade, earn more.<sup>177</sup> In 1984, 8.3% of Asian-Americans were in gifted programs and 1.6% in special education classes; comparable statistics for whites were 4.7% and 4.2%, respectively, though intelligence scores were similar for the two groups.<sup>178</sup> Some theorists attribute this phenomenon to hard work, persistence, concentration, deferred gratification, discipline, respect for authority, and family unity.<sup>179</sup>

Asian-Americans obtain maximum benefit from America's equal access and opportunity to education.<sup>180</sup> Possibly Asian-American students believe that with the right to education comes responsibility.<sup>181</sup> If students are to profit from the opportunities given them, they must be willing to work hard.<sup>182</sup> This means foregoing a night on the town with friends in

---

tional system may be affected by home and other social environment outside the classroom. . . . [T]his case at least suggests that minority race children shall not be subjected to and afforded quality education until history has been repealed and the fact that there was segregation be no longer a fact." *Id.* at 1089.

172. *Id.* at 1086. In the initial administration of the SSAT II, 78% of blacks *failed* one or more sections. *See* Debra P. v. Turlington, 474 F. Supp. 244, 248 (M.D. Fla. 1979) by 1983, 91% of blacks *passed* the SSAT II. *See* Debra P. v. Turlington, 564 F. Supp. 177, 187 (M.D. Fla. 1983).

173. Debra P. v. Turlington, 654 F.2d 1079, 1086 (5th Cir. 1981). *See also*, Note, *supra* note 10, at 879 (suggesting that a positive remedy requiring the schools to cure the students' educational deficiencies would have been more appropriate than an injunction).

174. Lee & Rong, *The Educational and Economic Achievement of Asian-Americans*, 88(5) THE ELEMENTARY SCH. J. 545 (1988).

175. *Id.*

176. *Id.* at 557.

177. *Id.* at 548.

178. *Id.* at 558-59.

179. *Id.* at 547, 554, 559.

180. *Id.*

181. *See* Calfee, *Standards, Evidence and Equity: Implications of the 1983 Debra P. Decision*, 2(4) EDUC. MEASUREMENT: ISSUES & PRAC. 11 (1983) ("students need to shape up").

182. *See generally*, Lee & Rong, *supra* note 174.

order to finish homework and get a good night's sleep.<sup>183</sup> It means getting to class on time, listening carefully in the classroom, *not* interfering or being disruptive when the teacher or other students are talking, completing and handing in homework on time, and refraining from participation in illegal activities.<sup>184</sup> Being responsible means being disciplined; it means putting one's education first in priority.<sup>185</sup> It also means *not* being lazy or wanting something for nothing.<sup>186</sup>

Some advocate that the schools are responsible for whether children learn.<sup>187</sup> Such advocates assert that we should stop blaming failure on the child's home environment.<sup>188</sup> But studies have shown that the home environment is one of the best predictors of school success;<sup>189</sup> thus, schools should not be blamed for inadequacies at home.<sup>190</sup> How can we expect the schools to be effective when students act irresponsibly and parents insist that school discipline infringes on their children's individual rights?<sup>191</sup>

### E. *The Rights of Other Students*

Concern does not stop with individual rights of a single student.<sup>192</sup> When a student fails academically and becomes disruptive, that student gets maximal protection of their individual rights.<sup>193</sup> Who protects the rights of the twenty other children while the disturbed child molests other children, makes loud obscene comments during the lesson, runs around the room throwing furniture, bangs his head against the wall, or destroys the work of other students, thus requiring the teacher to spend most of the day trying to keep the disruptive student from hurting himself or others.<sup>194</sup> When students are unable to profit from classroom instruction due to their emotional disturbance, is it fair to the other twenty children to

---

183. *Id.*

184. *Id.*

185. *Id.*

186. *Id.*

187. Snider, *Black Community Must Press for Reforms, Black Scholars Assert*, EDUC. WK., May 31, 1989, at 1, col. 2; at 22, col. 1.

188. *Id.* ("What we demand is this: that the schools shift their focus from the supposed deficiencies of the black child and the alleged inadequacies of black family life to the elimination of the barriers that stand in the way of academic success among these and other children.").

189. *Debra P. v. Turlington*, 564 F. Supp. 177, 188 (M.D. Fla. 1983).

190. See generally Gaddy, *High School Order and Academic Achievement*, 96 AM. J. OF EDUC. 496, 505 (1988); *Debra P. v. Turlington*, 654 F.2d 1079 (5th Cir. 1981) (Tjoflat, J., dissenting); *Debra P. v. Turlington*, 564 F. Supp. 177, 188 (M.D. Fla. 1983) (testimony of Dr. Barbara Lerner).

191. See generally, Lee & Rong, *supra* note 174, at 545, 560.

192. See Coons, *Shanker Vouchers — Is The Genie Out?*, 63 PHI DELTA KAPPAN 255 (1981); Lerner, *supra* note 9, at 252-53.

193. See *Goss v. Lopez*, 419 U.S. 565 (1975).

194. *Id.*



disrupt their learning also?<sup>195</sup> If the disruptive student is unable to learn due to his handicap and fails the graduation test, must all the other students be deprived of the opportunity to demonstrate their mastery of essential skills?<sup>196</sup> Do the rights of the emotionally disturbed child (or those of the minority child) take precedence over rights of the other children?<sup>197</sup> The collective individual rights of the twenty children who are attentive ought to be worth as much as the individual rights of one.<sup>198</sup> Protection of one student's rights should not result in concurrent infringement on the rights of many other students.<sup>199</sup>

### *F. The Lowest Common Denominator*

It is totally unrealistic to believe that all students can or will learn the same amount in a given period of time, unless the instructional objectives are reduced to the lowest common denominator.<sup>200</sup> If schools teach only what the weakest, most irresponsible student learns, the majority of students are slighted.<sup>201</sup> However, if tests are not used to find out what students know, the weaker students might receive no attention at all.<sup>202</sup> Standardized assessments are the means to determine if students have learned what they should.<sup>203</sup> But interpretation of standardized assessments must be tempered by the fact that teachers do not have sole

---

195. *Id.*

196. *Id.*

197. *Id.*

198. *Id.*

199. *Id.*

200. See dissent by Judge Hill in denying rehearing of the *Debra P.* case, *Debra P. v. Turlington*, 654 F.2d 1079, 1087 (5th Cir. 1981) (Tjoflat, J., dissenting).

201. See generally, Walsh, *Private Education Firms Discovering A Lucrative Market in Public Schools: Schools "Could Do Just What We Do if They Had the Same Resources,"* EDUC. WK., June 7, 1989, at 7. "When the courts placed an injunction against the high school graduation program for 4 years, student performance remained static. When the injunction was lifted, student performance immediately rose. In the end, if students are expected to perform, they will." Turlington, *How Testing is Changing Education in Florida*, 4(2) EDUC. MEASUREMENT: ISSUES & PRAC. 9, 10 (1985).

202. See generally, Rothman, *High-Stakes Testing: A District Ties Goals to Scores*, EDUC. WK., March 22, 1989, p. 1, col. 1, p. 8, col. 1.

203. Lerner, *A National Census of Educational Quality — What Is Needed?*, NASSP BULL., March 1987, p. 42. Note that "standardization test" means only assessment under uniform conditions for all students; it is *not* synonymous with the term "Multiple-choice test." Standardized assessments include performance measures such as essay writing, open-ended problem solving and skill demonstrations. Mehrens and Lehmann define uniform procedures as "the same fixed set of questions administered with the same set of directions and timing constraints, [with scoring procedures that are] carefully delineated and kept constant." W. MEHRENS & I. LEHMANN, *USING STANDARDIZED TESTS IN EDUCATION* 7 (4th ed. 1987).

control over whether children learn.<sup>204</sup> The student is also responsible for learning.<sup>205</sup>

### IX. Conclusion

Dr. Barbara Lerner speaks of the difference between feel-good-now self-esteem and self-esteem earned through the hard work of achieving standards.<sup>206</sup> Testing helps a state determine student progress as compared to national achievement and where more resources are needed.<sup>207</sup> Diploma tests may also ensure that students are given the opportunity to learn those skills essential to becoming a productive member of society.<sup>208</sup> A recent article in *Time Magazine* stated: "[T]he nation, once proud of its frontier individualism, has gradually adopted a no-risk mentality based on the belief that if anything bad happens, someone should be made to pay."<sup>209</sup> Let us ensure that it is *not* our children! *Let us teach our children to achieve success the old-fashioned way: by earning it!*

---

204. Gaddy, *High School Order and Academic Achievement*, 96 AM. J. OF EDUC. 496, 505 (1988).

205. *Id.*

206. Lerner, *Student Self-Esteem and Academic Excellence*, 52 THE EDUC. DIG. 32 (1986).

207. *See generally*, Lerner, *supra* note 206, at 1058, 1064.

208. *Id.* at 1059-60.

209. Church, *Sorry, Your Policy is Canceled*, TIME, Mar. 24, 1986, at 26.

