1998

Yearning for Lake Wobegon: The Quest for the Best Test at the Expense of the Best Education

Lisa Kelly
University of Washington

Follow this and additional works at: https://digitalcommons.law.uw.edu/faculty-articles

Part of the Education Law Commons

Recommended Citation

This Article is brought to you for free and open access by the Faculty Publications at UW Law Digital Commons. It has been accepted for inclusion in Articles by an authorized administrator of UW Law Digital Commons. For more information, please contact cnyberg@uw.edu.
YEARNING FOR LAKE WOBE GON:
THE QUEST FOR THE BEST TEST
AT THE EXPENSE OF THE
BEST EDUCATION

LISA KELLY*

That's the news from Lake Wobegon, where all the women are
strong, all the men are good-looking and all the children are above
average.

Garrison Keillor¹

INTRODUCTION

Every Saturday evening for many years, the audience has smiled
at the conclusion of Garrison Keillor's weekly report from Lake
Wobegon, the sleepy little Minnesota town that time forgot. Listeners
are amused by the insistence that all children, or at least all of the
children of the homogeneous community peopled by Weavers, Bun-
sens and Ingqvists, must be above average. The joke is buried in the
mathematical truth that we all know—not all can be above average
because the average covers the entire range of abilities that constitu-
tes us all. The humor also springs from our immediate recognition
of that American vanity that takes for granted that its children are, of
course, brighter than all others.

Nevertheless, our legislated educational policy and recent presi-
dential pronouncements have been driven by an engine of mandates
chasing after outcomes not too different from those yearned for in

*Professor of Law, West Virginia University, College of Law. Many thanks to Randy
Berkey, a school psychologist, who discussed the ideas in this piece with me for hours on end. I
also appreciate the able assistance of Brian Peterson, my research assistant, whose enthusiasm
for the project fueled his ability to track down many of the sources upon which I have relied.
Wiley Newbold also assisted in finalizing the citation in this piece and I am thankful for his
diligence. Like all worthwhile teaching and learning, this Article has been the product of a
collaborative process.

1. GARRISON KEILLOR, GOSPEL BIRDS AND OTHER STORIES OF LAKE WOBEGON (Minne-

41
Lake Wobegon. The executive branch, federal statutes and state plans have pushed for more and earlier testing to ensure that every school meets high standards. The results of this testing can be used for many explicit purposes: to reward schools where children meet the standards; to punish schools whose children don’t; to insist that children master certain skills by certain ages; and to retain them if they don’t. The wisdom of these explicit goals is rarely challenged in the political arena because of the Lake Wobegon effect; we all believe in the inherent good of a world where all children are pushed to be above average. Being soft on children and their school work may not be as politically suicidal as being soft on crime, but certainly few elected officials are willing to stand up and question these policies without fear of sounding as though they expect less from American children than they should. In the mainstream public debate, rarely are these goals examined more closely to see whether they comport with sound educational theory.

In addition to the explicit goals of measuring school outcomes and individual student success, these testing requirements also have effects, perhaps unintended, that should be considered carefully before we as a nation continue to hurtle down this road of more and earlier testing of elementary school children. These effects include incentives for teachers to sacrifice developmentally appropriate curricula in order to teach to tests and a deepening of racial and class differences in an educational system that sacrifices children of color and poor children to make the dream of Lake Wobegon come true for the children of privilege. Children of the poor and communities of color become the leaven that allows for the remainder of American children to rise above average.

This article first will outline the various tests or assessments. Next, recent federal and state mandates for standardized testing of elementary school children will be examined. Then, the educational literature will be reviewed to expose the dangers of testing, particularly in the early grades. Finally, I will urge that the Clinton administration, Congress, and the states step back from this manner of securing educational adequacy. These governmental policies are rooted in the stated intention of guaranteeing that all schools are doing right by all of the children. However, early testing fosters the opposite result—educational inequity through tracking, retention, and the early creation of a racial and class caste system. Furthermore, widespread testing of the type advocated by the federal government is
an expensive proposition. Given the negative effects of testing, I advocate the use of these funds in other ways to address children's real educational needs.

I. AN ASSESSMENT PRIMER

Today, the variety of available assessment or testing instruments is staggering. However, most assessments can be categorized as either norm-referenced or criteria-referenced. Norm-referenced tests have long been the most familiar tool for measuring educational success in the United States. Norm-based tests measure the student's performance against a sample of students used to set the norm. These assessments, typically multiple choice tests sold by commercial suppliers, are designed initially to produce curved results, with half of the students scoring above average and half below. Typical examples of norm-based tests are the *Iowa Test of Basic Skills*, the *Stanford Achievement Test* and the *Comprehensive Test of Basic Skills*.

One of the built-in problems with most norm-referenced tests is that the norm is established by using a national sample of students before the test is actually used in the market. At this early stage, perhaps test results truly do reflect a student's position in the national continuum. However, over time, as the instrument becomes more familiar and teachers begin to teach to prepare their students to take the test, the original norm is no longer reflective of the current range of student performance. Unless re-normed frequently, test results become inflated.

This inflationary effect, now known as the "Lake Wobegon effect," was discovered by a West Virginia physician, John B. Cannel, who treated children and their families. He knew from his experience that many children were experiencing depression and stress-related illnesses linked to their schooling. He also knew that West Virginia, one of the poorest states in the Union, had one of the highest illiteracy rates. Nevertheless, when the standardized test scores were announced, the children of West Virginia were found to be performing way above the national average. Dr. Cannel began checking other

---

states' results and learned that all of the thirty-two other states surveyed were basking in the pride of having above average children. It seemed that the impossible Lake Wobegon Utopia had been achieved. Dr. Cannel's further investigation revealed that the tests used were operating on norms that had been established some ten years prior to the time all states became safe havens for above-average children.\(^5\)

Norm-referenced tests embody the worst of two worlds. On the one hand, if accurately normed, the tests are set up to stigmatize half of the children as being below average. On the other hand, usually the norms are so old that test results at the statewide level provide a false feel-good impression that nearly all of the children are above average. Outdated, norm-referenced tests serve no real educational purpose at all but can function as excellent public relations tools for the states themselves.

Criteria-referenced tests, on the other hand, measure a student's performance against an unwavering set of curricular standards. These assessments measure a student's knowledge against the content which s/he is expected to possess, not against what other students may or may not know.\(^6\) Most Department of Education policymakers favor criteria-referenced tests over norm-based tests. They urge that content standards should be set high, assessments should be aligned with these challenging standards, and educators and students should be held accountable for the results.\(^7\)

A third, but rarely used, form of assessment is by portfolio review.\(^8\) Under the portfolio review system, samples of the student's work throughout the year are collected and reviewed by scorers outside of the school to determine the student's evolving knowledge and abilities.\(^9\) This form of assessment provides probably the most

---

5. See Fiske, supra note 4, at 119-20.
7. See id. at 19.
9. What Are Promising Ways to Assess Student Learning?, Improving America's School: A News. on Issues in School Reform (U.S. Dep't of Educ.), Spring 1996. See also
accurate picture of the school’s curriculum content as well as the student’s demonstrated abilities. Among its disadvantages, however, is the subjectivity of the review, the cost of review, and, for those advocates of standardization, an inability to state the results in quantifiable, comparative terms.\textsuperscript{10}

There are two basic forms that assessment questions can take, no matter whether the tests themselves are norm-referenced or criterion-referenced. These two forms are performance-based or multiple choice. Under performance-based testing, the student is required to demonstrate directly what s/he can do.\textsuperscript{11} For example, instead of filling in a series of multiple choice bubbles at the end of a math word problem, the student will be required to answer the question showing his or her work. The choices and guess work are gone; the student must use the tools in his or her head. Another example is in writing.\textsuperscript{12} Multiple choice tests may be able to test for knowledge of the rules of standard grammar or they may present a student with choices for the best missing sentence in a particular paragraph.\textsuperscript{13} They cannot, however, test the student’s ability to compose, organize and write his or her own paragraph on a given or self-selected subject. Performance-based assessments place the child in the position of demonstrating his or her abilities in these previously untested skills. Some test instruments include both multiple choice and performance-based testing.

Performance-based testing is more costly than the traditional multiple choice tests. A recent study by the General Accounting Office found that a national multiple-choice achievement test would cost $42 million. By contrast, a purely performance-based test would cost almost five times as much—$209 million—to administer and score nationally. About one-fourth of the $209 million dollars would


\textsuperscript{11} See Fiske, supra note 4, at 115-16, 133-34.

\textsuperscript{12} Both the Education Week and the Council of Chief State School Officers reports break writing assessment out as its own unique form of testing, separate and apart from performance testing. See High Standards, supra note 8, at 34; Trends, supra note 8, at 35. However, to the extent that both performance based assessment and the more specific writing assessment require the student to demonstrate his or her knowledge of the subject matter tested, I use “performance assessment” as the more general term to refer to both.

\textsuperscript{13} See Fiske, supra note 4, at 117-19.
need to be expended on professional development for teachers and scorers.  

Finally, assessments vary as to their purposes. Some instruments are geared to measure individual student achievement while others are meant to be used to measure regional or school performance. Assessments of student achievement can be used for class placement purposes, decisions on whether to retain or promote a student to the next grade level or to design remedial programs for the individual child. Assessment of regional or school performance can be used to punish or reward schools and teachers based upon their classes' performance, to make personnel decisions, and to target funds for remedial programs. They can also be used to hold schools accountable to the parents and the taxpaying public. Sometimes educational policymakers in government have multiple expectations for what one testing instrument can achieve. Out of concern for cost, they dream of an instrument that can be all things for all people. However, testing experts warn against such shortcuts. Tests that truly can assess individual student strengths and weaknesses are much more detailed and nuanced than those geared toward discovering a school or district's ranking statewide or nationally.

Standardized tests can take many forms and serve many purposes. Next, we will attempt to discern what is being advocated and implemented on the national and statewide levels.

II. LEGISLATIVE POLICY SUPPORTING STANDARDIZED TESTING

A. The Presidential Push for Early Standardized Testing of Elementary School Children

As President Clinton began his second term in office, he set forth his desire to encourage the establishment of national standards to ensure that all children learn what it takes "to succeed in the knowledge economy of the 21st century." The President spoke of the need

---

17. President William J. Clinton, Address on the State of the Union 4 (Feb. 4, 1997) (transcript on file with author) [hereinafter Clinton, Address].
for educational excellence with the urgency and commitment of a general planning for battle. He referred to the need for standards as a matter of national security and called upon a citizen army of tutors to secure the goal that every child will be able to read independently by the age of eight.

President Clinton lamented that forty percent of our 8-year-olds cannot read on their own, and pledged that his administration would "lead an effort over the next two years to develop national tests of student achievement in reading and math." He urged every state to "test every 4th grader in reading and every 8th grader in math to make sure that these standards are met."

This testing would depart from the usual goal of assessing school or regional performance and would be aimed instead at measuring individual student performance. The President summarized the usefulness of such testing to high standards as follows:

Good tests will show us who needs help, what changes in teaching to make, and which schools need to improve. They can help us to end social promotion. For no child should move from grade school to junior high, or junior high to high school until he or she is ready.

The President has been reiterating the same or similar educational goals in addresses to state legislatures. In an address to the Michigan legislature, the President elaborated further upon his plans to establish "high standards, high expectations, and high levels of accountability." Again he urged the testing of every fourth-grader in reading and every eighth-grader in math by 1999. In this address,

18. Id. at 4.
19. Id.
20. See id. Testing is not the only method by which the President seeks to improve student performance to meet high standards. As noted in the text, he has proposed that college students serve as Americorps volunteers to help children learn to read independently. Under the President's budget, Head Start funding will be increased, the formation of charter schools to pioneer new methods of learning will be supported, and additional funding will be available to rebuild deteriorating school infrastructures. See id. at 4-5.

The focus of this article upon testing should in no way be taken as a criticism of these other initiatives, some of which may indeed be very helpful or even essential to continued educational success. The deteriorating physical structure of our schools has long been documented and lamented by such prominent educators as Jonathan Kozol. See JONATHAN KOZOL, SAVAGE INEQUALITIES (1991). During the President's first term, Congress passed legislation geared to providing federal money to schools most in need of repair. See 20 U.S.C.A. §§ 8501-8676 (West Supp. 1997).

22. President William J. Clinton, Remarks by the President to the Joint Session of the Michigan Legislature 8 (Mar. 6, 1997) (transcript on file with author) [hereinafter Clinton, Michigan Remarks].
he promised that the Department of Education will support the development of a new test to measure every student by “high and widely accepted standards.”

The President assured the Michigan legislature that the federal government will not mandate that these tests be administered in every state in the Union. However, he did state his intention “to create a climate in which no one can say no; in which it’s voluntary but you’re ashamed if you don’t give your kids the chance to do this.” In Michigan, Clinton reiterated the goal that every 8-year old should be able to read independently and issued the optimistic pronouncement that “90 percent of the children in this country plus—99 percent of them—can learn what they need to know to succeed and triumph in the modern world.”

President Clinton gave the same speech to the North Carolina legislature while the Vice President was busy delivering a virtually identical message to the California legislature. Clearly, testing and high educational standards are a top priority on the Clinton-Gore agenda for this second term.

The President continues to travel the country delivering his stump speech in support of early testing of children. In a recent trip to West Virginia for a town hall meeting, the President was more detailed about the type of test he advocates for fourth-and eighth-graders. He was clear that he preferred criteria-referenced tests as opposed to norm-based testing. He also provided examples of the types of questions he supports, one of which was performance-based and the other, multiple choice. On the reading test, he asked a fourth grader from the audience to read a passage from Charlotte’s Web. He then showed the question that would be asked following the passage. The child would formulate his or her response based upon his or

23. Id. at 9.
24. Id.
25. See id. at 10.
26. Id. at 7.
28. See Vice President Albert Gore, Remarks by the Vice President to the California Legislature (Mar. 13, 1997) (transcript on file with author).
her own understanding of the passage. There were no multiple choice options. The response would then be scored based upon the detail of the answer. The sample math problem was a typical word problem with multiple choice responses. Hence, the President appears to be advocating a criteria-referenced test with mixed performance and multiple choice questions to be used to assess individual student performance. Notably, in the examples he gave, the performance-based test question would have been given to the fourth-grader while the eighth-grader would have been given the multiple choice question. In effect, the fourth grader was subjected to the more rigorous form of testing.

On April 29, 1997, Secretary of Education Richard W. Riley appeared before the House Subcommittee on Early Childhood, Youth, and Families to defend the administration’s commitment to placing the development of this national test on the fast track, without prior Congressional approval. The test’s development costs during the 1997-1998 years will be approximately $22 million, according to Secretary Riley. This amount will come out of the Department’s $40 million fund for improvement in education. The test is slated to be administered for the first time in 1999.

The President also placed educational issues on the front burner during his first term. In response to the call for curriculum-content standards, the President followed through with former President Bush’s project of commissioning the Department of Education to assemble specialists to establish a national curriculum. However, the search for a national curriculum proved to be unsuccessful. Some

32. See Paul Gagnon, What Should Children Learn?, ATLANTIC MONTHLY, Dec. 1995, at 65, 68. Clinton was instrumental in the early stages of the educational reform movement initiated during the Bush administration. In his capacity as the head of the National Governor’s Association, he catapulted to national prominence for his leadership on education reform issues during the National Governor’s Summit on Education which met in September of 1989 and again at the National Governor’s Association meeting in February of 1990. See David S. Broder & David Hoffman, Bush, Governor’s Chart Ambitious School Goals, WASH. Post, Sept. 28, 1989, at A4; Kenneth J. Cooper, As Bush Hails ‘New Era,’ Educators Voice Concern, WASH. Post, Feb. 27, 1990, at A4.

As governor of Arkansas, he put the state on the map of educational reform by placing Hillary Rodham Clinton in charge of a statewide educational reform effort that netted major changes in testing for teacher and student competency, a longer school day and year, and mandatory kindergarten. See Kenneth J. Garcia, Arkansas Under Clinton, S.F. CHRON., Oct. 9, 1992, at A1; Robert McCord, Don’t Underestimate Hillary Rodham Clinton, USA TODAY, Jan. 27, 1993, at 11A; Jim Nichols, Arkansas Joins Coalition Targeting School Reform, ARK.
projects were defunded for nonperformance, while others were derailed by controversy.\textsuperscript{33} The proposed world-history curriculum faced the biggest public storm, as ideological wars erupted over such issues as whether Western icons like the Magna Carta received too little attention in the voluminous history curriculum at the expense of the attention paid to the Mongol empire and "sub-Saharan Africa."\textsuperscript{34} Eventually, amid the sound and fury of the canon debates, the national curriculum was jettisoned in favor of the creation of financial incentives for states to develop their own curriculum standards. This push for state standards was enacted in the Goals 2000 legislation.\textsuperscript{35}

Clinton appears to have learned from his first term that national consensus on educational issues is hard to come by, and in his second term, he has moved quickly to advocate the same voluntary, incentives-based model for states to comply with an agenda of earlier testing to challenging standards. In fact, the Goals 2000 and other first-term federal statutory reforms also included the seeds for the type of early standardized testing advocated by the President during his second term.

Next, we will look at the federal legislation already in place regarding testing. After reviewing the national test which has been in place since 1969, we will turn below to those federal statutes which provide the current catalyst for state testing reforms.

B. FEDERAL LEGISLATIVE SUPPORT FOR TESTING IN ELEMENTARY SCHOOL

1. \textit{The National Assessment of Educational Progress}

The National Assessment of Educational Progress ("NAEP") was developed in 1969 to test student achievement nationally. Since 1988, it has been used for state-by-state comparison purposes only. The NAEP is a national criteria-referenced assessment that looks to whether the tested individual has attained a basic, proficient or

\textsuperscript{33} \textit{See} Gagnon, \textit{supra} note 32, at 68.

\textsuperscript{34} \textit{See id.} at 74-75.

advanced level in reading, writing, math, science, and history/geogra-
phy. The test includes both multiple choice and performance-based
questions.\footnote{36. See \textit{Student Achievement, supra} note 3, at 27.}

Although the NAEP is a creature of federal statute,\footnote{37. See 20 U.S.C.A. § 9010(d)(1) (West Supp. 1997).} participation by the states is voluntary and not all states participate in the testing for all subjects at all levels.\footnote{38. See \textit{Student Achievement, supra} note 3, at 27.} To guarantee that the test will only be used for comparative purposes, the NAEP Commission is expressly prohibited from releasing any personally identifiable data with regard to test subjects.\footnote{39. See id. § 9010(b)(1)(A).} It is to be administered at least once every two years on students at ages nine, thirteen and seventeen; and in grades four, eight and twelve in both public and private schools.\footnote{40. See id. § 9010(b)(1).} Because the test is not concerned with individual student performance, only a representa-
tive sample of students from each locale is tested.\footnote{41. See Millicent Lawton, \textit{Broaden Achievement Definitions for NAEP, Panel Urges, Educ. Wk. on the Web} (Apr. 9, 1997) <http://www.edweek.org/ew/vol-16/28naep.h16>.}

Even though the NAEP test does include some performance-based questions, the test has been criticized by education scholars as being still too focused upon paper-and-pencil skills. The National Academy of Education at Stanford University has suggested that in the future the NAEP should be broadened to test for problem-solving, interpretation, and performance in groups.\footnote{42. See id. §§ 6301-6514.} We will explore further the state results on the NAEP and comparisons of those results with state-formulated assessments in Part II. C. 2. below.

2. Goals 2000: Educate America Act and the Strengthening and Improvement of Elementary and Secondary Schools Act

Many of the goals advocated by the President now are the same or very similar to those already set forth in the Goals 2000: Educate America Act\footnote{43. 20 U.S.C.A. §§ 5801-6084 (West Supp. 1997).} and the closely aligned Strengthening and Improvement of Elementary and Secondary Schools Act,\footnote{44. Id. §§ 6301-6514.} often referred to as Title I, which targets educational reforms at disadvantaged children.

Goals 2000 sets forth national education goals as the foundation upon which the remainder of the lengthy statute is based. Among
these goals are: 1) "by the year 2000, all children in America will start school ready to learn"45, 2) "by the year 2000, all students will leave grades 4, 8 and 12 having demonstrated competency over challenging subject matter including English, mathematics, science, foreign languages, civics and government, economics, arts, history and geography, and every school in America will ensure that all students learn to use their minds well so they may be prepared for responsible citizenship, further learning, and productive employment in our Nation's modern economy"46, and 3) "by the year 2000, United States students will be the first in the world in mathematics and science achievement."47

Just how educators are to achieve these lofty goals is left to the states and local educational agencies to devise. States may apply to the federal government for funds to establish content standards and state improvement plans geared to meeting those standards.48 As a part of the plan, which must be submitted at the time of application for federal funds, the states must also outline their plans for student assessments.49

Goals 2000 describes criteria-referenced tests which must be aligned to curricula. The statute has language which suggests that performance-based measures are encouraged. The statute does not describe specifically for what purposes these tests should be done—whether for individual student assessment or regional comparisons. However, the language is broad enough to include measuring individual student performance as at least one of the purposes of assessment.

Specifically, Goals 2000 requires that state assessments be nondiscriminatory, valid, reliable, and monitored. Further, the assessments must: 1) be aligned with the State's content standards; 2) use multiple measures of student performance; 3) be able to be used by students with diverse learning needs; 4) allow for accommodations and adaptations for students with those diverse learning needs; 5) be consistent with nationally recognized professional and technical standards for such assessments; 6) provide the state with coherent information about student attainment of the standards; and 7) support effective curriculum and instruction.50

45. Id. § 5812(1)(A).
46. Id. § 5812(3)(A).
47. Id. § 5812(5)(A).
48. See id. § 5885.
49. See id. § 5885(b)(2).
50. See id. § 5886(c)(1)(B).
The Goals Panel was created in part to work with and assist the states with the technical support needed to develop such instruments, in particular with regard to early childhood assessments used to gauge school readiness. As of the fall of 1996, Goals 2000 had provided $1,986,000 in grants to eight states and a consortium of twenty-two others to develop assessments. These grants set the states free to develop their own assessments.

While the Goals Panel is to review the standards and assessments and report to the President on the work of the states, the criteria that the Department of Education uses to review them are broad and flexible. Specifically, a state improvement plan must: 1) hold reasonable promise of helping all students achieve at the high levels called for under the Goals 2000 Act; 2) reflect widespread commitment within the State and; 3) allow local schools, local educational agencies and communities the flexibility to implement local improvement plans that reflect local needs.

As of the fall of 1996, forty-five states had received second-year funding from Goals 2000, and nineteen states had submitted comprehensive improvement plans to receive third-year funding. Since 1994, $1,270,270,000 has been allocated to the states from the federal government to support these requests. This total represents an escalating amount each year, and the fiscal year 1998 promises to be no exception. The projected amount for 1998 will be $610,000,000.

Even more money is available to the states under the Act entitled Strengthening and Improvement of Elementary and Secondary Schools, or Title I. The general purpose of Title I is to close the gap between the educational performance of disadvantaged children and children of more affluent communities. The total budget for these

---

51. See id. §§ 5823-5827.
52. Goals 2000 Activity Across the Nation, GOALS 2000: A PROGRESS REPORT (U.S. Dep't of Educ.), Fall 1996, at 4, 5. The eight states receiving individual grants are Delaware, Maryland, Michigan, Minnesota, North Carolina, North Dakota, Oregon, and Pennsylvania. See id. at 5.
55. The specific outlay for each fiscal year from 1994 to 1997 is as follows: 1) FY 1994 - $92,400,000; 2) FY 1995 - $361,870,000; 3) FY 1996 - $339,700,000; 4) FY 1997 - $476,000,000. For a breakdown by state and territory and by year, see Goals 2000: Building on a Decade of Reform, GOALS 2000: A PROGRESS REPORT (U.S. Dep't of Educ.), Fall 1996, at 2.
56. Interview with Jay McClain, Goals 2000 Specialist, Dep't of Educ. (May 19, 1997).
57. See 20 U.S.C.A. § 6301 (West Supp. 1997). The portion of this Act dedicated to disadvantaged children is more popularly known as "Title I."
programs for next year is projected to be $7.1 billion dollars. The funding formula is implemented on a county-by-county basis using child-poverty counts, and every state in the Union has been participating for its qualifying counties.

During the first term of the Clinton administration, Title I was amended to bring the same demand for challenging standards to disadvantaged students as it did for all students under Goals 2000. Under this program, now states must submit plans to the federal government for increasing targeted students' performance based upon challenging standards developed under Goals 2000 or similar testing systems. Again, assessments are a key part of the plan. These assessments have all of the requirements placed upon the Goals 2000 assessments and then some. The statute specifically requires that each student have his or her own individualized results; that the data be disaggregated by gender, race, ethnic, English language proficiency, migrant, and disability status; and that the students be tested in those areas in which the state has adopted content standards some time during the following: 1) grades three through five; 2) grades six through nine; and 3) grades ten through twelve.

Because the current and proposed federal legislation place so much power in the hands of the states, we must look at state trends to see the true impact that testing spurred by federal incentives has had on elementary school children. The next section will review these trends and look to the work of a few states that typify certain assessment approaches.

C. State Assessments of Elementary School Children

1. Assessments Used by the States

The types, number, and purposes of assessments employed by the states is in a constant state of flux, as states try out new methods under

59. See id. at 35.
62. See id. § 6311(b)(3).
the encouragement of Goals 2000 and amended Title I. All but four states have some form of statewide assessment.\textsuperscript{63}

Of the four states that do not have a testing program up and running, one, Minnesota,\textsuperscript{64} currently has its assessment system in the development stage. Only two, Iowa and Nebraska, continue to resist the idea of vesting the state with control over local school districts.\textsuperscript{65} Iowa’s effort to create statewide standards and assessments was halted in 1993 by conservative Christian groups and others, who characterized standards and testing aligned with them as an effort by Big Brother to force a “politically correct” curriculum down local school districts’ throats.\textsuperscript{66} Nevertheless, Iowa does participate in a number of national norm-referenced tests which indicate that eighty-eight percent of Iowa elementary and middle schools score above the national median in basic skills.\textsuperscript{67} Nebraska also maintains a strong commitment to local control, and a spirit of resistance to statewide standards and assessments thrives among the people. Local school districts develop their own standards; no statewide assessments exist.\textsuperscript{68}

Of the remaining states that already do have statewide assessments in place, the majority use a combination of types of assessments, which are constantly changing. Twenty-nine states relied upon norm-referenced tests for at least a portion of their testing; and five of them used norm-referenced tests exclusively. Only four allowed for any assessment by portfolio review. Thirty-four included a criteria-referenced test as one measure, thirteen of them without the use of any norm-referenced tests. Twenty-three states included a performance component. Thirty-six also tested their students’ writing abilities.\textsuperscript{69}

However, states continue to change their minds about testing instruments. Most of the debate centers around whether tests should

\textsuperscript{63} The Council of Chief State School Officers Report shows four states with no statewide testing: Colorado, Iowa, Minnesota, and Nebraska. The Council reports that Wyoming and Missouri sample students statewide. Wyoming uses a performance-based test, and Missouri uses a multiple choice criteria-referenced test with a writing assessment component. Finally, California leaves the decision to test up to local school districts. See TRENDS, supra note 8, at 35-36.

\textsuperscript{64} See High Standards, supra note 8, at 34.

\textsuperscript{65} See TRENDS, supra note 8, at 35-36.

\textsuperscript{66} See Bess Keller, Going Against the Grain, EDUC. WK., Jan. 22, 1997, (Supplement), at 110.

\textsuperscript{67} See id. at 108.


\textsuperscript{69} See TRENDS, supra note 8, at 35-36.
be performance-based or use the familiar multiple choice model. The desire to fall back upon multiple choice, norm-referenced tests, which are cheaper and make for better public relations, remains strong.\(^7\)

For example, Vermont, which had been among the first states to abandon multiple-choice tests in favor of portfolio review, will be returning standardized tests to the mix. Vermont remains committed to portfolios, however, as the best tools for individual student evaluation, but norm-referenced multiple-choice tests will return for the purpose of comparing schools. The new Vermont plan, to begin in the 1997-1998 academic year, calls for a reading test in second grade; English-language-arts and math tests in the fourth, eighth, and tenth grades; science tests in the sixth and eleventh grades; and social studies tests in the sixth, ninth, and eleventh grades.\(^71\) Kentucky, long an innovator in curriculum and assessment, is also considering a proposal to include more multiple choice questions.

Some states nix performance-based tests before they are ever used. When the Indiana legislature considered the $100 million to $50 million price tag and subjective scoring problems associated with a performance-based exam, it rejected the proposed test in favor of a largely multiple choice exam.\(^72\) Idaho has continued to expand its use of norm-referenced, multiple-choice tests. Now grades three through eleven will take the *Iowa Test of Basic Skills*.*\(^73\) In California, the California Learning Assessment System was abandoned for multiple choice tests to be given to all students in grades two through ten in reading, spelling, writing, and math. However, local school districts could opt out of the testing program. In a move to boost the flagging enthusiasm of the many California school districts, each district participating in this testing program is being offered $5.00 per student taking the test.*\(^74\)

West Virginia, the home of the Lake Wobegon effect, dabbled with performance-based testing, but has moved recently to norm-referenced, multiple choice tests administered as early as kindergarten.\(^75\)

\(^72\) See Lawton, supra note 70.
\(^73\) See id.
\(^75\) West Virginia will use the Stanford 9, a product of the Harcourt Brace Publishing Company, to yield percentile ranks for school to school and county to county comparisons for the
When asked why West Virginia was placing so much emphasis upon norm-referenced standardized test instruments, State Superintendent Henry Marockie pointed out the $100,000 West Virginia has spent on assessment compared to Kentucky's nearly $29 million and answered, "These nationally normed tests have done a fabulous job of picking out the textbooks what is really important, and we're all convinced that in the end, we will be just as well off and better, without spending all the money." Perhaps these normed tests do such a good job of picking out the important parts of the text books because they are published by the text book company, Harcourt Brace. Tying one's measure of success to an inexpensive standardized test that uses a particular standard text book company's offerings certainly hamstrings curricular reform. True to its Lake Wobegon origins, West Virginia also has established as an educational goal that "student performance on national measures of student performance will equal or exceed national averages and the performance of students falling in the lowest quartile will improve by fifty percent."
Clearly, states are struggling with the tensions around the purpose and form of assessments. What we have now is a crazy quilt of different standards and mixtures of assessments nationally, and it is a quilt that seems to be redone annually. It is probably too early to tell where the majority of the states will settle, but at present the trend appears to be toward a mixture of assessment tools with a heavy dose of the multiple-choice norm-referenced tests.\textsuperscript{79} Standardized norm-referenced tests are familiar, relatively cheap and, if they are not re-normed regularly, soothing to the soul. Their results are much more comforting than the results of criteria-referenced tests, as a review of the NAEP results will show.

2. State Results on the National Assessment of Educational Progress Test

The NAEP is a national test tied to challenging standards. What have the results been? Forty-five states participated in the 1996 eighth grade math assessments, and in 1994 fourth grade reading assessments were conducted in thirty-nine states.\textsuperscript{80} Only 28% of those fourth grade students who took the NAEP reading test in 1994 scored at the proficient level or higher. The highest scoring state was Maine with 41% scoring at or above proficiency; the lowest scoring state was Louisiana, with only 15%. In Louisiana, 60% of its tested fourth graders scored below the basic level in reading.

The eighth grade math results in 1992 were even lower. Nationally, only 21% scored at or above proficiency. Iowa and Minnesota were tied for first place with 31%. Mississippi came in last with only 6% scoring at or above the proficient level and 67% scoring below the basic level.\textsuperscript{81}

An interesting picture emerges when the NAEP results are compared with state assessments. For instance, only 15% of Louisiana’s fourth grade students were reading at the proficient level according to the NAEP. However, according to state assessments using state standards, 88% were reading at proficiency levels. States from all over the country are experiencing the same effect. In Wisconsin, the NAEP found only 35% of the fourth graders reading at the proficient level, but according to state assessments, 88% have again met the standard.

\textsuperscript{80} See Student Achievement, supra note 3, at 27.
\textsuperscript{81} See id. at 24, 30.
reading level. In Connecticut, the differential is ten percentage points—38% according to the NAEP and 48% according to the state standards.\footnote{U.S. Dep't of Educ. et al., \textit{State NAEP Scores for 4th Grade Reading Compared to States' Own Assessments} (visited July 31, 1997) \langle\text{http://www.ed.gov/Speeches/04-1997/970429.html}\rangle.}

Many possible explanations for the disparity between NAEP and state assessment results immediately suggest themselves. It could be the Lake Wobegon syndrome at work through states using outdated standardized tests badly in need of renorming. Or it could be that the teachers are being driven to teach to the state assessments, and to ignore a different curriculum suggested by the NAEP. It could be that the NAEP standards are just too challenging or not developmentally appropriate.

Another factor will be explored in greater depth below, but should not be ignored here. Perhaps the NAEP assessment is culturally biased. The states with the best scores ironically reflect more closely the ethnicity and economic stability of our imaginary Lake Wobegon. Those with the lowest scores have high African American populations.\footnote{In Iowa, only 7% of its public school students are children of color and only 13% are poor. On the 4th grade NAEP reading test only 7% of Iowa's African-American children and 16% of the Hispanic children scored at the proficient level as compared to the 37% of white 4th graders. See Keller, \textit{supra} note 66, at 108, 110.}

Similarly, Minnesota, the homestate of Lake Wobegon, is known for its homogeneous population and low child poverty rates. To the extent that more diversity is being achieved in the Twin Cities, great disparities can be found in the NAEP results for African-American, Hispanic, and Native American children when compared to the white children. In the fourth grade reading NAEP test, only 9% of the African American children, 21% of the Hispanic children, and 15% of the Native American children scored at or above proficient, compared to the 35% of white children who did. See Ann Bradley, \textit{Standards vs. Control: Minnesota's Independent-Minded Districts Make Reform Efforts an Uphill Battle}, \textit{Educ. Wk.}, Jan. 22, 1997, (Supplement), at 139, 141-42.

Ninety-seven percent of Maine's students are white. There are so few African-American children there that statisticians do not even do break-outs of NAEP test results for them. Results for Hispanic children are included in the statistics, and they show that on the fourth grade reading test only 25% of Hispanic children scored at or above proficient compared to 42% of the white children. See, Ann Bradley, \textit{Building Support: Though Maine's Students Perform at the Top, Educators Remain Busy Pushing Reforms}, \textit{Educ. Wk.}, Jan. 22, 1997, (Supplement), at 122, 124-25.

By contrast to the above homogeneous islands of relative prosperity, 49% of Louisiana's student population is comprised of children of color and 35% of its children live in poverty. See Millicent Lawton, \textit{Hoping to Climb: Public Education Has Not Been a Priority in Louisiana, and Educators Will Have to Struggle to Change That}, \textit{Educ. Wk.}, Jan. 22, 1997, (Supplement), at 117. Mississippi's student population is 52% children of color and has a child poverty rate of 33%.
repercussions for communities of color and the economically disadvantaged. These disadvantages will be explored further in Part III.B.2., below.

III. THE PEDAGOGICAL DANGERS OF ASSESSMENT AND ELEMENTARY SCHOOL CHILDREN

With all of the statutory and monetary incentives for more and earlier testing in place, little attention has been paid to whether such widespread use and preoccupation with assessments are pedagogically sound. What follows will question some of the explicit goals and examine some of the negative unintended effects of the assessment programs that states are currently employing.

A. QUESTIONING THE EXPLICIT GOALS

1. The Use of Standardized Testing to stop “Social Promotion”

President Clinton has repeatedly stated that the use of assessments for individual school children can stop the evil of social promotion. The flip side of this coin is, of course, that children who do not measure up to the challenging standards assessed by the new criteria-referenced tests should be retained. But is retention a pedagogically sound practice? Does it help children to learn?

Study after study has found retention to be of questionable educational benefit to children. In fact, most studies have found retention to be a harmful practice. Retained students, when compared to similarly performing children who were promoted, did not benefit from the additional year in grade; in fact, they were far more likely to drop out of school once they reached high school. Retention in the early grades has been shown to result in an educational career characterized by low-level tracking and repeat non-promotions. So widely known and accepted are the negative effects of retention that the Best Practices in School Psychology III, formulated by the National Association

---

See Julie A. Miller, Delta Dawn: In the Campaign to Improve Education, Mississippi Has a Greater Challenge, EDUC. WK., Jan. 22, 1997, (Supplement), at 143.

84. See supra Part II.A.

of School Psychologists, warns against viewing retention as the answer to a child’s academic troubles.\footnote{86} Not only does the mere repetition of the same curriculum fail to address the student’s real problems and needs, but it brings with it an undeniable stigma. Everyone recognizes the negative connotations of the common parlance for retention. Who would want to be the student who was “held back,” “had to repeat a grade” or worse yet, simply “flunked?” One study found that children rated retention behind only blindness and a parent’s death as the most stressful experiences imaginable. Another found that retained students viewed their retention as “flunking” and as a punishment. Retained students suffer from learned helplessness in which they believe that they inherently are unable to succeed, and explain away any successes as having nothing to do with their ability but instead credit them to the kindness of others or luck.\footnote{87}

Some believe that, as long as retention occurs in the early grades, it is not as harmful to the child’s self-image. The extra year is believed to help the late-birthday child or give the immature child an opportunity to catch up. However, again the research points to the contrary. Longitudinal studies show that any gains that may have been made by the child repeating first grade, for example, are soon dissipated if they exist at all. Similarly situated children who were not retained in the first grade catch up to their grade level peers by grade four while the retained children are more likely to experience poorer academic and social functioning throughout the elementary grades.\footnote{88}

Retention disproportionately impacts African American and Latino students, particularly those in poor urban districts. A National Educational Longitudinal Study reported in 1988 that by the time all students reach the end of eighth grade, almost one of five have repeated a grade. However, more than one out of three low-income students will have experienced retention. Grade retention rates for African American and Latino students are three times as high as they are for white students. African American students and Latino students are retained earlier and more often than white students. This

\footnote{86} NATIONAL ASS’N OF SCHOOL PSYCHOLOGISTS, BEST PRACTICES IN SCHOOL PSYCHOLOGY III, at 413-19 (Alex Thomas & Jeff Grimes eds., 1995). \footnote{87} See id. at 415. \footnote{88} See id.
large group of retained children of color reappear in the same statistical pool that reflects the disproportionately high dropout rate for students of color.\textsuperscript{89}

Finally, retention costs money. Based upon current retention rates and average per pupil costs, we spend $10 billion nationally to pay for the extra year of what appears to be at best useless and at worst harmful schooling for children who are retained.\textsuperscript{90}

There are alternatives to retention that have proven to be more successful and less damaging to a child’s self-concept. Merely repeating the same curriculum with the same delivery system does not address the child’s real needs. More promising alternatives include cooperative learning, direct instruction, individualized instruction, peer tutoring, and special programming such as after-school tutoring, remedial reading or math classes, and even summer school.\textsuperscript{91}

Sometimes the problem is not with the child, but with the teacher or the curriculum as it is organized and delivered. Consideration should be given to expanding classroom teachers’ capabilities to deal with the needs of students at differing skill levels. Schools and classrooms may need to be reorganized to assure that each child can develop at his or her pace in a supportive environment.\textsuperscript{92} Alternatives to the rigid grade system should also be contemplated. Transition classrooms, or what is sometimes called split-grades (where children can operate at different grade levels in the same classroom), reduce stigma and maximize a child’s ability to work at his or her own pace. And, of course, open non-graded or individualized instruction classrooms would present the greatest degree of flexibility.\textsuperscript{93} Recognition must be had of the fact that young children in particular learn and grow at their own speed. Neither pathology nor failure need attach to a child who, in common parlance, is a late bloomer. The importance of a developmentally appropriate curriculum will be stressed in Part III.B.1 below.

\textsuperscript{89} See Keenan & Wheelock, \textit{supra} note 85, at 2.


\textsuperscript{92} See id. at 27.

\textsuperscript{93} See id. at 36.
Of course, the press for assessment based upon challenging standards only throws fuel on the already dangerous and costly fire of retention. Where we set the bar, i.e., at the achievement of the basic or proficient level, becomes critical, and can worsen a retention problem that already smolders and for some groups rages out of control. Given recent performance on the NAEP, for example, it would be madness to begin retaining the seventy-two percent of fourth graders nationally who failed to achieve the proficient level on the reading test.\textsuperscript{94} Even if promotion were conditioned upon achievement of the basic level of performance, that would still leave forty-one percent of the fourth graders behind.\textsuperscript{95} Surely, even this result cannot be acceptable.

2. The Reward and Punishment of Schools based upon Outcomes of Student Testing as a Means of Fostering Educational Accountability

The President has argued that his proposed tests would help to target unsuccessful schools. Most of the recent history of testing has been geared toward competition between states, school districts, and individual schools for the highest scores. Even those states that had moved away from the norm-referenced, multiple choice tests typically used for comparative assessments are restoring them to their battery of tests to be able to engage in comparisons of schools and districts within the state. This competition is advocated in the name of accountability and across-the-board improvement of education.

For example, Kentucky has long been regarded as one of the most innovative and daring educational reform states. Among the measures used by Kentucky is the reward-punishment system for school performance. When a school performs below standards, the school is given two years to meet them. If the school does not improve, the state considers it a “school in crisis” at which time a team of education specialists takes over the school. Faculty and administrators of the school in crisis are subject to transfer or dismissal. If, on the other hand, the school exceeds the threshold standards, financial rewards flow to it from the state—to be disposed of as the faculty sees fit.\textsuperscript{96}

\textsuperscript{94} See supra Part II.C.2.
\textsuperscript{96} See Lonnie Harp, Nowhere to Go But Up: Kentucky's Landmark Education Reform-Law Has Created a Rare Opportunity, Educ. Wk. Jan. 22, 1997, (Supplement), at 114; Lois
When closely studied, however, it becomes clear that high test scores do not always reflect school quality. The test-takers are not the faculty, but children. Children come to school with already developed skills and abilities, influenced by home, varying modes of expression and availability of resources. Typically, schools do not have a random sprinkling of children from different socio-economic and cultural backgrounds. Rather, schools tend to concentrate children of similar backgrounds in one building or region in a district. Therefore, the high test results of one school or region may merely reflect the access to resources and an immersion in the culture valued by the test rather than the school’s quality.

Some schools, conscious of the fact that certain types of students are more likely to have more desirable test results than others, attempt to enhance their test scores by controlling which students take the tests. Magnet schools can set certain criteria for admission that

---


97. For example, more than two-thirds of African American and Hispanic children attend schools in which the enrollment is predominantly African American and/or Hispanic. Lynn Olson, Examining Race and Demography, Educ. Wk., Jan. 22, 1997, (Supplement), at 10, 11. Rates of continuing educational segregation vary from state to state. Of the thirty-two states containing 98.2% of the United States’ African American population, Illinois and New York top the list for segregated schools. In Illinois, 83.2% of African American children attend predominantly black schools. In New York, 80.8% attend predominantly black schools. See Andrew Hacker, Two Nations: Black and White, Separate, Hostile, Unequal 162-63 (1992).

The Supreme Court’s recent decisions in the area of school desegregation reflect a retreat from the integration ideal and a capitulation to white flight in the guise of private choice. Consequently, segregated schools will likely continue to be an American institution. See Missouri v. Jenkins, 515 U.S. 70 (1995) (holding that orders designed to attract nonminority students from outside the school district into the school district was beyond the scope of the violation; that requiring across-the-board salary increases for teachers and staff to create desegregative attractiveness was beyond the scope of the court’s remedial authority; and that testing the achievement of unitary status by looking at whether students in district are at or below national norms was inappropriate); Freeman v. Pitts, 503 U.S. 467 (1992) (holding that the district court has authority to relinquish supervision and control over school district in incremental stages before full compliance with desegregation remedy has been achieved in every area of school operations); Board of Educ. v. Dowell, 498 U.S. 237 (1991) (holding that formerly segregated school districts may be released from court-ordered busing even if some segregation persists so long as all practicable steps to eliminate the vestiges of discrimination have been taken); Milliken v. Bradley, 418 U.S. 717 (1974) (preventing busing across city and county lines even if resulting school systems are predominantly white and black).

For a vivid narrative account of the continuing segregation of our schools by color and class, see Kozol, supra note 20.


99. For critiques of magnet schools as a remedy to racial segregation, see Kimberly C. West, A Desegregation Tool that Backfired: Magnet Schools and Classroom Segregation, 103 Yale L.J. 2567 (1994) (exposing racial classroom segregation within “integrated” magnet
will enhance the likelihood of attracting desirable students. A dynamic can also develop in some districts which allow students to transfer freely between schools. In such a district, high-scoring schools may place pressures on parents of children who perform less well to move their children to another school.100

Furthermore, the competition for the highest test scores can sometimes produce detrimental results in and of itself, not the least of which is the unseemly specter of teachers and administrators cheating on tests to improve or maintain high scores.101 In addition to cheating, perhaps a more common effect of the reward and punishment of competitive testing can be to place powerful incentives on schools not only to admit or keep the highest achieving students, but also to identify and place those with lower test results in special education classes where their test scores will no longer be counted.102 Inappropriate placement in special education can have lifelong consequences for children, who are then tracked into lower level courses of study.103 Placing undue weight on test scores also increases the likelihood of retention in the hope that students will score better in a grade lower than their placement.104

Children of color experience these detrimental effects disproportionately to their white counterparts. They are overrepresented in special education, disproportionately retained, and score lower on
standardized tests as a group. Increasing the importance of performance on standardized tests can only work to compound the educational obstacles that African American and Latino children face.\footnote{While African American students represent 16% of all public school students, they make up nearly 40% of those who are classed as mentally retarded, disabled, or otherwise deficient. \textit{Hacker, supra} note 97, at 164; \textit{Keenan & Wheelock, supra} note 85, at 2. \textit{See also Jeannie Oakes, Keeping Track: How Schools Structure Inequality} (1985).}

The competition for high test scores also has a negative impact upon the teaching resources available to lower-scoring schools. Poor children and children of color are less likely to be in classes with teachers who at least minored in their teaching fields than their white, middle class counterparts.\footnote{Specifically, in the 1990-91 school year, 42% of high school math classes taught to poor children and children of color were taught by those who were math majors. In high schools where children of color made up fewer than 15% of enrollment, 69% of the math classes were taught by math majors. \textit{See Olson, supra} note 97, at 11.} The higher scoring schools may attract the better teachers, who elect not to subject themselves to the punishments visited upon teachers in less successful schools.\footnote{See \textit{Darling-Hammond, supra} note 98, at 40.} Instead of having the intended effect of making all schools accountable to all children, standardized testing for comparative purposes can actually have the opposite result of retrenching the educational inequities already in place for children who score the lowest on the very tests allegedly put in place to help them.

\section*{B. Examining the Unintended Effects}

Standards and assessments impact upon the type of curriculum that children are subjected to beginning with kindergarten. As demonstrated above, teachers are motivated by rewards and punishments to teach to tests. But does teaching to tests, particularly in the early elementary grades, really comport with an appropriate curriculum for very young children?

Many of the negative consequences of the standards and assessments movement have already been shown to hit communities of color the hardest. But are there other dangers inherent in the tests themselves that make them potentially biased for children of color? These two questions will be examined in the following sections.
1. Sacrificing Developmentally Appropriate Curriculum to Teach to Tests

The term "Kindergarten" was first used by German educator Friedrich Froebel, and translates literally into "the child's garden." Under the original concept of kindergarten, children were given gifts that they then used to explore and manipulate the world around them. The first gift is a ball, then a cube, yet another a puzzle composed of cubes. And through manipulation of these gifts the child began to learn important concepts and social skills. A few schools in the United States today still follow this German model kindergarten curriculum. Germany is one of the countries that American education reformers repute to be ahead of us today.

In 1962 in the United States, kindergartners were expected to learn how to play well with others, how to explore the world around them with fingerpaints and music, and how to accomplish such complex small motor skills as lacing and tying a shoe. Perhaps, through the art of song, the kindergartner would also learn the alphabet. And at story time, which usually came right before nap time, the seeds of sitting and listening to the wonders of language would be planted. It was a milk and cookies time, kindergarten was. Some readers may remember it that way.

Today, kindergartners are expected to learn the difference between a word and a sentence, divide words into syllables, recognize sound-letter relationships for consonants, two-letter consonant-blends and digraphs, and write with good posture, holding paper and pencil correctly and using uniform spacing between words and letters. They must learn to capitalize the months of the year, days of the week, and the first letter of people's names. They are expected to read national


110. See Gagnon, supra note 32, at 71.

111. These descriptions come from this author's recollection of kindergarten. The lacing of the big wooden shoe was the most daunting task. If you need an outside source to confirm the simplicity of kindergarten during this era, see Lorrie A. Shepard & Mary Lee Smith, Escalating Academic Demand in Kindergarten: Counterproductive Policies, 89 ELEMENTARY SCHOOL J. 135-36 (1988).
and international authors and be able to identify the correct uses of punctuation, such as periods, question marks, and exclamation points. They are to learn how to interpret charts and graphs and, of course, become familiar with test-taking strategies. In math, they must know how to count forward to 100 and backward from 10, both with and without objects to assist them; they must recognize patterns of counting by fives and tens. They must estimate the size of an object and compare objects with respect to length and weight. They must run screaming from the room in search of milk and cookies.

It is no wonder that readiness for school has become an issue. No longer are children permitted to ease into a formal educational milieu. According to Goals 2000, all children must be ready to begin school, and not just any school, but a school with a challenging curriculum. Consequently, many districts have begun administering tests to determine whether a child should be admitted to kindergarten. These tests, once again, have a disproportionate impact upon economically disadvantaged children and children of color, children who also often stand most in need of the milk and cookies experience to bring them into contact with a positive educational experience.

Another consequence of these tough standards and testing is that parents are electing to wait to send their children to school. This means that children whose parents can afford to “red shirt” them into an extra year of pre-school enter that year with advantages over their less affluent peers to whom they are later compared. The curriculum becomes geared to the highest performing, older children, and again the less advantaged child runs the risk of being labeled and left in the dust of a challenging kindergarten curriculum. In yet another effort to make sure children are ready for school, states continue to raise the

112. See Henry R. Marockie, Instructional Goals and Objectives for West Virginia Schools 21-24 (1996). These standards and objectives were formulated by the State Superintendent of Schools as part of the Goals 2000 state plan to set challenging standards. This represents just a few highlights from three pages of a six-page kindergarten curriculum which includes language arts, mathematics, social studies, and science.


115. See id.
minimum age for kindergarten enrollment, thereby depriving disadvantaged children not involved in Head Start programs of exposure to early education.116

The fast track only speeds up after kindergarten. This push for early learning geared toward rigid curricula and tests has produced what educators have dubbed the “hurried child syndrome.”117 The hurried child syndrome is the result of America's assembly line approach to education in which all children are expected to learn the same, usually superficial, things at the same pace. The rapidity of learning does not allow for the development of curiosity, deep knowledge, or critical thinking. Worst of all, the wearying pace does not allow for a true love of continued learning to grow. Tomes containing K-3 curricula stifle the ability of truly creative teachers to instill a zest for learning that all children should be greeted with at the first moments of schooling. And testing is the straight-jacket that forces each teacher to comply. Ironically, one of the frequent complaints of school reformers, echoed by President Clinton, is that those countries outperforming us educationally construct their curricula so that children spend longer periods of time mastering an in-depth knowledge of fewer subjects, while we hurry through, giving children only a superficial understanding at best of too many ideas.118 The tools we have chosen to foster educational reform beget the very results we bemoan.

The fact is that young children learn at different speeds and in different ways. Very few young children learn well in a rigid classroom setting, in which they are expected to sit up straight, be quiet, and inordinately preoccupy themselves with whether they are holding their pencils correctly. Young children perform best in the context of a curriculum which is developmentally appropriate. They flourish when exposed to a curriculum that focuses on making the school ready for the child and not the other way around.119

116. In 1958, most districts required that kindergartners be five years old by December or January 1st. In 1988, the common cut-off was October 1. See Shepard & Smith, supra note 111, at 138. Today, the more common cut-off is September 1. See id.


Developmentally appropriate education recognizes that young children learn by manipulating their environment. The process for learning between the ages of three to at least six is to explore, manipulate, and then conceptualize. The learning of older children gradually moves from these concrete learning forms to more symbolic and derived methods of learning. And each child varies in the pace at which they move from the concrete to the symbolic.\footnote{Howard Gardner, Harvard Professor of Education, has advocated that this period of exploration of materials ought to continue until at least the age of seven.} Developmentally appropriate education also embraces learning through play.\footnote{In general, it is a program that emphasizes "direct experience, adult warmth, concrete materials, child-initiated activity, and social interaction."} In harmony with these general directives, developmentally appropriate education recognizes that a child’s learning needs are not solely academic. As Dr. James Comer, Maurice Falk Professor of Child Psychiatry at the Yale Child Study Center, has noted, “School success depends on adequate child development along six critical pathways: 1) speech and language; 2) physical; 3) moral; 4) social-interactive; 5) psychological-emotional; and 6) academic-intellectual.”\footnote{These pathways are interconnected and mutually reinforcing. Failure to attend to all of these developmental needs, Dr. Comer suggests, “lies at the heart of American education problems.”}

One example of the importance of a developmentally appropriate curriculum can be found in a comparative study of kindergarten...
teachers who were found to have a high retention rate with those whose retention rates were low. The analysis of the data generated by this study indicated that teachers who had high retention rates used more developmentally inappropriate teaching tools such as formal reading instruction with a higher emphasis on the taking of tests than their developmentally appropriate and successful peers.126

The criticality of a developmentally appropriate curriculum starts at kindergarten, but it does not end there. The National Association for the Education of Young Children ("NAEYC") has recognized and described inappropriate practices for teaching in all primary grades. Among those inappropriate practices is the construction of curriculum narrowly by focusing upon the intellectual domain valued solely by the acquisition of discrete technical skills. Another relevant inappropriate practice is measuring children's worth "by how well they conform to group expectations, such as their ability to read at grade level and their performance on standardized tests."127 Finally, according to the NAEYC, children should not be evaluated using norm-referenced tests.128 Neither an educational system driven by norm-referenced tests (preferred by some states) nor criteria-referenced tests aligned to "challenging" but developmentally inappropriate standards (the darling of federal policymakers) bodes well for true education reform.

2. Fostering Racial Bias in Education

As already demonstrated, poor children and children of color experience the negative consequences described above most sharply. They are retained disproportionately; their schools already provide fewer resources than the schools of privileged white schools;129 they lose in the competition for the highest test scores, thereby lessening their status among teachers who quite naturally desire positive reinforcement over public ridicule; they suffer from not being admitted into kindergarten, because they are found to be disproportionately not ready; and they are tracked disproportionately into special education classes.

126. See Zepeda, supra note 123, at 74.
127. NAEYC, Guidelines, supra note 122, at 67.
128. See id.
The federal statutes pay lip service to the fact that assessments must be non-discriminatory, but the criteria that Goals 2000 specialists use to review the state’s plans do not include criteria to reject funding if the assessment is in fact discriminatory; nor is any study required to determine the non-discriminatory nature of states’ tests before funding is disbursed.\textsuperscript{130}

Indeed, the Department of Education acknowledges that racial minorities score lower on traditional norm-referenced assessments, and that these assessments are in large part responsible for those students’ disproportionate representation in remedial and lower level tracks. In addition, the Department of Education also admits that research indicates that the gap between white students and students of color will likely widen with the use of performance-based testing. The reason, the Department hypothesizes, is that minority students are disproportionately placed in these lower level classes and are not exposed, therefore, to the more challenging material that would allow them to acquire the types of skills being tested in performance-based exams.\textsuperscript{131} The downward spiral continues as all of the negative consequences of testing converge upon the children who stand in most need of educational reform.

Some may argue that the testing instrument is merely the messenger bearing the bad news, and that the bad news needs to be heard so that it can be corrected. We need to know that poor children and children of color are not receiving the same educational quality as white, more affluent children. The tests are not racist or classist; they just bring us news of a racist and classist educational system. Undoubtedly, it is true that testing cannot be blamed for all of the evils of racism and classism inherent in our educational system and in the nation at large. This argument, however, ignores two important variables: 1) the extent to which testing has contributed to creating the circumstances in which disadvantaged children are retained, tracked, and excluded from school; and 2) the extent to which the tests themselves may be culturally biased.

The existence of this first variable has already been demonstrated above. The second variable is always more hotly debated. However,


\textsuperscript{131} What the Research Says About Student Assessments, Improving America’s Schools: A Newslet. on Issues in School Reform (U.S. Dep’t of Educ.), Spring 1996.
an example from the now infamous Ebonics debate may serve to demonstrate the subtleties of cultural bias in testing.

In 1973, Missouri psychologist Robert L. Williams coined the term “Ebonics” at the conclusion of a conference funded by the National Institute of Mental Health. The conference examined the cognitive and linguistic development of African American children. Ebonics was used to describe the African American linguistic system which represented the patterns and codes that black children brought to school with them, but which were not accepted and were often denigrated by the school’s different set of linguistic patterns and codes. A study followed in which African American kindergartners were asked questions typically found in standardized tests and re-asked them using a test recoded with a sensitivity for Ebonics. The differences in the questions themselves were sometimes slight, but the effect upon the test results were notable. For instance, in one question, the tester showed each child a picture of squirrels in different actions and asked the child to point to the squirrel that was beginning to climb the tree. Some of the children were able to do so, but others were stymied. When the question was reformulated using the phrases “starting to” or “fixing to” in place of “beginning to,” however, those children who were previously stumped answered easily and correctly.

This type of study sheds light not only on the ways that the language of standardized tests can be culturally biased, but also offers insight into how performance-based tests might also suffer from cultural bias. Consider for a moment the Ebonics study findings in the context of the performance-based reading comprehension tests advocated by President Clinton. In such a test, the student must respond by writing his or her understanding of the text as opposed to choosing

132. The Ebonics program most recently debated was adopted by the Oakland, California School District as an effort to deal with the problems reflected by low grade point averages and high suspension rates among its African American student body. Ebonics was to be used as a method of moving students from the use of Black English to the more standard English that would enable them to succeed in standard testing situations and to deepen their interest in and understanding of the curriculum. The controversy ignited over many aspects of the program: whether and how explicitly to value the language of Ebonics that at its heart reveals a great deal about the tensions over assimilation and subordination that swirl in the midst of many racial controversies; the ill-advised wording adopted at one point during the controversy that Ebonics was “genetically based”; and whether Ebonics was a separate language, a dialect, or more derisively, just plain “slang.” See Renee Sanchez, Ebonics—Without the Emotion, Wash. Post, Jan. 13, 1997, at 29; Patricia J. Williams, The Hidden Meanings of 'Black English,' N.Y. Times, Dec. 29, 1996, at E9.

from a list of multiple choice answers. A scorer, who is not sensitive to the particular linguistic patterns that a child brings to school, may score down the answer if it uses an idiom with which the scorer is not familiar. The child may comprehend the story well; the problem may lie in the scorer's comprehension. This type of cultural bias has been studied in classroom interactions between white teachers and young black students. These classroom observations have found that white teachers often misinterpret as dysfunctional communicative development the narrative speech patterns of African-American children in which many anecdotes are told together in a string.\textsuperscript{134} The subjectivity inherent in the scoring of many performance-based activities makes these tests vulnerable to the same forms of cultural bias as the multiple choice tests.

Assessments have had a checkered history as applied to African Americans. The most rudimentary "intelligence tests" of the eighteenth century used craniometry to determine that Africans were an unintelligent species more closely aligned with apes than white humans.\textsuperscript{135} In contemporary society, testing is used to disproportionately place African Americans in special education.\textsuperscript{136} Indeed, today's \textit{Bell Curve}\textsuperscript{137} controversies rest upon the continued faith placed in tests by those who argue for genetically pre-determined intellectual superiority among the upper classes of American society. Skepticism of any testing instrument that can so early and so radically determine an African American child's future would seem to be properly placed, particularly when the results have generally been as devastating as they have in the realm of education.

Testing, therefore, also has the more generalized consequence of reinforcing notions of white racial superiority. Tests help white skin privilege to thrive, and perpetuate the badges of inferiority worn by children of color. Critical race theory would characterize this is as a


\textsuperscript{136} See \textit{id. at 156-158; Carl Milofsky, Testers and Testing} 35-37 (1989); Joseph P. Shapiro et al., \textit{Separate and Unequal}, U.S. News & World Rep., Dec. 13, 1993, at 46, 48-49. For an examination of the ways that the learning disabilities movement with its attendant specialized testing, as distinguished from special education, developed as a white middle class reaction to avoid having their children placed in predominantly black classrooms, see Gerald Coles, \textit{The Learning Mystique: A Critical Look at "Learning Disabilities"} xii-xiii, 205-07 (1987).

\textsuperscript{137} Richard J. Herrnstein & Charles Murray, \textit{The Bell Curve: Intelligence and Class Structure in American Life} (1994); see \textit{also The Bell Curve Wars: Race, Intelligence, and the Future of America} (Steven Fraser ed., 1995).
benefit to white parents, who can feel confident that their children are above average because "others" will be permanently entrenched at the bottom. Indeed, the standards and assessments movement has grown out of the support of a white upper and middle class constituency, with little collaboration between those in the African American community active in school desegregation or other reform movements supported by communities of color.138

Professor Derrick Bell's white interest convergence theory offers a possible explanation for the persistence of the Lake Wobegon effect in the education reform movement. It also perhaps holds the key to genuine education reform. The white interest convergence theory holds that unless and until whites perceive some benefit to themselves in change, there is little chance that a system which supports white superiority will cede power to alternatives.139 It is true that the current standards and assessments movement does support white superiority. However, it also poses dangers, particularly in the early grades, not just to children of color and the poor, but to all who are subject to the hurried child syndrome and retention. Educating white parents about these dangers is critical if an effective challenge to these programs is to be mounted.

IV. CONCLUSION

The stated intention of educational reform is to provide all American children with an in-depth education, comparable to our foreign peers, an education that will enable them to compete in the global economy of the twenty-first century. The chosen method to achieve this goal is to hold schools accountable for achieving educational excellence through testing. However, assessments can actually frustrate real reforms and do serious harm.

At least two of the stated purposes of the standards and assessment movement actively damage children—by encouraging retention and by fostering competition between schools that drain the most needy of already limited resources. Poor children and children of color experience these negative consequences most keenly. They also must deal with the stigmatizing effects of tests that are culturally biased.

The damage done to all children by assessments is particularly acute in the early elementary grades, when children are first exposed to the educational system. The first years of school set the tone and sometimes the track that children will follow throughout the rest of their academic careers. If the assessments are not geared toward a developmentally appropriate curriculum, and if the assessments themselves are not developmentally appropriate, all children suffer.

Assessments are also very expensive, particularly those that are performance-based and criteria-referenced. Alternative, cheaper instruments have limited utility. Norm-referenced, multiple choice tests say little about what individual children actually know. They also work directly against the type of critical thinking and deep understanding that educational reformers say are the main goals of their movement.

These costly programs leave little money to implement them in effective ways. Teachers are not trained, consulted with, or provided resources to assist them with the new curricula, and unmanageably large class sizes make impossible the implementation of truly beneficial educational reforms that could help all children. Similarly, the critical roles of family and community participation have been overlooked by a movement that has developed tunnel-vision around measuring educational outcomes by assessments.

Even though the policy behind these statutes and state plans is to achieve parity with, indeed superiority over, our international counterparts, we are approaching educational reform in a truly outdated American way. Like the old factories that have shut down around us, we follow an assembly line approach to education, in which productivity is measured by the volume of the output—how many items in a voluminous curricular checklist we can prove that our children know. Meanwhile, we bemoan that our children learn their materials only superficially, while the children of our worldwide counterparts have developed the critical thinking skills and deep knowledge that will enable them to master the details of disciplines on their own.

We need to turn away from investing our faith and resources so heavily in testing, particularly for the elementary grades where testing

142. See Comer, supra note 124, at 113.
can do such long-lasting harm, and begin to look at what steps will really accomplish our educational goals. Priorities must be set. Money is needed to shrink class sizes so that real teaching can take place.\textsuperscript{143} Infrastructures must be rebuilt so that school buildings are habitable and accessible.\textsuperscript{144} Our system of financing education through local property taxes long has been the subject of criticism for the inequities it produces.\textsuperscript{145} It multiplies injustice to test all schools and students as though they have access to the same materials, resources, and infrastructure when the bleak truth is that wide disparities exist, sometimes between neighboring districts.

Standards may be necessary for some broad sense of accountability, but they should be standards that are general and follow the dictates of a developmentally appropriate curriculum cognizant of the many areas in which young children need to develop. If it is educational accountability that we desire, we should look to those adults we hold responsible, not the young children who have no choice but to trust them. To the extent that we hold individual teachers accountable for achieving these standards, we can use alternative methods such as requiring teachers to submit their plans for implementing developmentally appropriate curricula. We could expend dollars currently dedicated to testing on professional development and consultation.

\begin{footnotes}
\item[143.] In President Clinton's 1998 State of the Union Address, he announced that as a part of his 1998 balanced budget proposal, he would fund the hiring of 100,000 new teachers “to reduce class size in the 1st, 2nd, and 3rd grades to an average of 18 students a class, all across America.” The White House, Office of the Press Secretary, \textit{State of the Union Address by the President} (Jan. 27, 1998) \texttt{<http://www.pub.whitehouse.gov/uri-res/12R?pdf://oma.eop.gov.us/1998/01/27/11.test.l>}.
\item[144.] President Clinton also proposed that a portion of the 1998 budget be dedicated to modernizing or building 5,000 schools. \textit{See id.}
\item[145.] \textit{See Kozol, supra note 20; Symposium, Investing in Our Children’s Future: School Finance Reform in the 90’s, 28 HARV. J. ON LEGIS. 293 (1991). In San Antonio Independent School District v. Rodriguez, 411 U.S. 1 (1973), the United States Supreme Court rejected a Texas student’s challenge of the state’s property tax-based school financing system and the inequities it produced. In so doing, the Court ruled that education was not a fundamental right under the federal Constitution and that poor students in poorly funded districts did not constitute a suspect class entitled to strict judicial scrutiny. The Supreme Court’s refusal to find any federal constitutional right to educational equity has led plaintiffs to turn to their state constitutions where some have been successful in restructuring school financing using alternative statewide systems of revenue raising and disbursement. See Marcia A. Brown-Thunberg, \textit{Raising Revenue for an Adequate Education in New Hampshire}, 20 VT. L. REV. 1001 (1996); Peter Enrich, \textit{Leaving Equality Behind: New Directions in School Finance Reform}, 48 VAND. L. REV. 101 (1995); Bradley W. Joondeph, \textit{The Good, The Bad, and the Ugly: An Empirical Analysis of Litigation-Prompted School Finance Reform}, 35 SANTA CLARA L. REV. 763 (1995); Jennifer M. Palmer, \textit{Education Funding: Equality Versus Quality—Must New York’s Children Choose?}, 58 ALB. L. REV. 917 (1995); Suzanne M. Steinke, \textit{The Exception to the Rule: Wisconsin’s Fundamental Right to Education and Public School Financing}, 1995 WIS. L. REV. 1387.}
\end{footnotes}
with individual teachers if necessary, including training in meeting the diverse needs of all students.

In the later grades, we need to review teacher qualifications and change them to require that teachers have degrees in the subject that they teach. Only twelve states require teachers to possess college course work other than in education for licensure.\(^\text{146}\) If we truly want students to be delving deeply into particular subject matter, it would seem logical to require that the teachers attain the knowledge required for that level of mastery. Raising the credentials of teachers would also require raising their compensation, but the costs and benefits of securing the most qualified teachers should be measured against the costs and benefits of developing the best test. In this cost-benefit analysis, if the goal is good teaching, then it would seem that attracting the most knowledgeable teachers should win.

Horror stories always spread of the teacher who can fill out the paper work or obtain the degree, but who sleeps through the class.\(^\text{147}\) If we are serious about the need for accountability at the extreme end of the continuum, then consideration should be given to auditing teacher performance by outside reviewers and for taking serious disciplinary action against teachers who are not delivering services, no matter how pretty their paperwork is. Perhaps requiring that teachers maintain portfolios of their students work subject to random review would also satisfy the need to ascertain the real workings of the classroom.

Placing accountability directly upon teachers, rather than attempting to measure student performance, may avoid many of the negative repercussions of testing discussed in this article. Of course, in shifting one's gaze directly to teachers, respect would have to exist for teachers as professionals with individual teaching styles directed to their unique student populations. A one-size-fits-all approach will not work for teaching, anymore than it will for learning. However, in the early elementary grades a teacher's plans and execution should be measured against a developmentally appropriate yard stick. Finally, unless and until our system for financing education is restructured, recognition would also have to be had that teachers work in settings

\(^{146}\) The twelve states are California, Colorado, Connecticut, Iowa, Massachusetts, Missouri, New Hampshire, New Jersey, New York, Tennessee, Texas, and Virginia. See Teachers Who Have the Knowledge and Skills to Teach to Higher Standards, EDUC. WK., Jan. 22, 1997, (Supplement), at 42, 43.

\(^{147}\) See MILOFSKY, supra note 137, at 4.
with varying access to resources. It would be unfair to create a system which penalized teachers for the poverty of their districts.

Testing may be the stuff of which good politics and public relations is made, but it is not the stuff of good teaching. True education reform requires that the federal government, states, schools, and teachers let go of the vanity that makes them yearn for Lake Wobegon and the best test, and look instead for real education reforms that will make a positive difference in the lives of all children.