



Additional copies of this report can be ordered for \$10.50 each (prepaid) from:

Policy Information Center  
Mail Stop 19-R  
Educational Testing Service  
Rosedale Road  
Princeton, NJ 08541-0001

(609) 734-5949  
e-mail: [pic@ets.org](mailto:pic@ets.org)

Copies can be downloaded for free at:  
**[www.ets.org/research/pic](http://www.ets.org/research/pic)**

Copyright © 2009 by Educational Testing Service. All rights reserved. ETS, the ETS logo and LISTENING. LEARNING. LEADING. are registered trademarks of Educational Testing Service (ETS). AP, Advanced Placement and Pacesetter are registered trademarks of the College Board. Educational Testing Service is an Affirmative Action/Equal Opportunity Employer. 11215

June 2009

# Table of Contents

|  |    |
|--|----|
| Preface.....   | 2  |
| Acknowledgments.....   | 2  |
| Introduction.....  | 3  |
| Where We Have Been .....   | 5  |
| The Current Conversation About National Standards .....  | 7  |
| What Does Setting National Standards Mean? .....   | 9  |
| Some Recent Developments.....  | 10 |
| Recognizing the Wide Variation in the U.S. Education System .....                                  | 13 |
| Variation in Content.....  | 13 |
| Variation in Performance Standards .....   | 15 |
| Variation in Student Achievement .....   | 19 |
| The Importance of Widespread Variation in Achievement to the National Standards Discussion....     | 23 |
| Who Bears Responsibility for Meeting Standards?.....   | 25 |
| Facing Choices, Risks, and Difficulties .....  | 28 |
| Relationships Between the Call for National Standards and Existing Test-based Accountability ..... | 28 |
| Assessing the Benefits of Local and State Responsibility .....                                     | 28 |
| Recognizing Continuing Conflicts.....  | 29 |
| Creating a National Standards System .....   | 31 |
| What to Do?.....   | 32 |
| Learning from Experience.....  | 32 |
| Finding What’s Feasible and Useful.....  | 33 |
| Curriculum Standards and “Standardization” .....   | 34 |
| An Alternative to Consider .....   | 35 |
| The One-Size-Fits-All Question .....   | 36 |
| Standards for Course Titles?.....  | 36 |
| Getting More Out of the Current National Efforts.....  | 38 |
| Concluding Comments.....   | 40 |

## Preface

Establishing national standards for what should be taught in schools along with a test to measure if the standards have been met have been debated and tried to varying extents over the last quarter century. Each time, the investment in high-level commissions and commissioned development work has ended in inaction. Controversy over the effectiveness of the No Child Left Behind Act (the federal government’s major influencer of state education behavior), the persistence of achievement gaps, and concern about our international competitiveness have resulted in new efforts to establish national or common standards. Some advocate for national standards to be developed and implemented outside of the federal government, while others are pushing for federal legislation.

While the public discussions have been ongoing for quite a while, they gained new momentum since the presidential election of 2008 and the anticipation of a reauthorization of the No Child Left Behind Act.

This report, authored by Paul E. Barton, is written with the objective of increasing our understanding of the history, the facts, the

choices, the risks, and the possibilities that are relevant in coming to a decision about establishing national standards in a nation that has built its education system from the bottom up and has prized local control of the schools. Barton makes clear that the report’s purpose is to inform, rather than oppose or advocate a particular course of action. It also is clear that he thinks the considerations are many, and that complexities abound. Nevertheless, he offers a number of approaches that might be considered for increasing commonality in what is taught, and presents examples and experiences that are being tried across the nation, and might be built upon in moving in this direction. While the report is not a “yes” or a “no” about uniform national standards, the clear message is that anyone who wants to make a sound and reasoned judgment on the question needs to do much homework first. This report will help with that.

Michael T. Nettles  
Senior Vice President  
Policy Evaluation and Research Center

## Acknowledgments

A number of people provided helpful reviews of this report. They include Richard Coley, Christopher Cross, Emerson Elliott, Drew Gitomer, John Jennings, Mark Musick, Michael Petrilli, W. James Popham, Richard Rothstein,

and Charles Smith. Not all reviewers would necessarily agree with all of the conclusions drawn. Janet Levy and Eileen Kerrigan were the editors. Marita Gray designed the cover and Sally Acquaviva provided desktop publishing services.

# Introduction

The education reform movement did not begin with a call for national or federal action, but for action among schools, districts, and states. That call, sounded by recommendations in the 1983 *A Nation At Risk* report, came from the National Commission on Education Excellence (NCEE) — at a time when President Reagan advocated the abolition of the new Department of Education. Against this backdrop, a strong national education policy appeared unlikely.

Although much transpired at lower levels to implement the NCEE report's recommendations, underneath the surface brewed a desire for action at the top. So began a period of high-level action caught in a start-stop loop, but never gaining clear-cut momentum.

By the beginning of the 21st century, strong action was taken at the federal level. This came in the form of the No Child Left Behind (NCLB) Act, which specified what states *had* to do regarding raising student achievement and improving the quality of teaching. While NCLB now has an unknown future, it has illuminated the issue of national education standards. The act has renewed in many a thirst for such standards and rekindled in others an aversion to them.

This report will discuss issues involved in the debate over whether the United States should have national education standards, what must be considered in creating such standards, what problems must be addressed, and what trade-offs might be required among conflicting objectives. The hope is that posing questions and providing relevant information will help clarify thinking about this topic.

The first section provides a short summary of developments in education over the past couple of decades. The current movement toward common

standards has not suddenly appeared, but comes with a history of starts and stops — evidence that a clear and lasting consensus has not yet emerged.

Next follows information about who has been saying what about bringing national standards into existence; then, what advocates and detractors mean when they talk about standards. Often, their meanings differ.

Many people have indeed given serious thought to the standards issue, and collaborations have developed to find commonality. This report summarizes progress on that front.

The “Recognizing Variation” section picks up on one of the hot topics of today. Underlying the effort for standards and commonality is an understanding of the significant variations nationwide in curriculum, content, and assessment quality. This report addresses three distinct kinds of variation: *content/curriculum*, *performance standards*, and *student achievement*, the latter of which is typically expressed in terms of a cut-score on a test.

Discussions of standards have not always identified which set of variations are being addressed, thereby not fully identifying what needs fixing. Comprehending the complexity of all three variations is important to the standards discussion, as differences within each type can be extraordinary.

Finally, this report advances some of the choices that need to be made, risks that those choices will entail, and difficulties that will be involved in achieving common standards. Many of these issues arise from the country's long history of building its education system from the ground up and at the local level, as well as from the great diversity of local and state populations and their economies.

An important question is: How should the country get started? How can it create an entity that sets standards *and* gains widespread acceptance? This report presents some thoughts, particularly on the topic of finding a process that could lead to standards reaching a recognized legitimacy.

Providing a clear set of answers is beyond the capabilities of this author, but this report explores a few avenues for moving toward greater commonality. Included are thoughts on gaining more from the National Assessment of Educational Progress (NAEP), which has seen expanded use over the past 25 years.

Overall, the hope of this report is a modest one: that readers will increase their knowledge on the topic and will gain a clearer understanding of what is possible and desirable.

## Where We Have Been

For historical context on education standards, four references are recommended: *National Standards in American Education: A Citizen's Guide*, by Diane Ravitch (Brookings Institution Press, 1995); *Why National Standards and Tests?*, by John F. Jennings (Sage Publications, 1998); *Political Education: National Policy Comes of Age*, by Christopher Cross (Teachers College Press, 2004); and “Standards Based Reform: A Powerful Idea Unmoored,” by Lauren Resnick, et al., in *Improving on No Child Left Behind*, edited by Richard D. Kahlenberg (Century Foundation Press, 2008). These sources better equip readers to consider what might be avoided, what might be repeated, and what might be the prospects for new national standards efforts in the future.

The first salvo on national standards did not come from the government. In 1989, the National Council of Teachers of Mathematics (NCTM) published a set of standards for teaching mathematics, based on consensus from many teachers and mathematics experts. Former Colorado Governor Roy Romer, who headed the National Education Goals Panel, said these standards exemplified what needed to be done in other subject areas.

Indeed, the NCTM standards did serve as a model for similar efforts in other subjects, initiated by and funded under the leadership of Diane Ravitch, an assistant Secretary of Education in the administration of President George Herbert Walker Bush. President Bush first set the education agenda when, in 1989, he convened the nation's governors at an education summit in Charlottesville, Va. Emerging from this conference was a set of national goals to be reached by 2000 — thus marking an era of collaboration.

Ravitch's efforts, meanwhile, concerned rigorous content rather than standardized tests and accountability, with voluntary standards created in science, history, geography, foreign languages, the arts, English, and civics in 1991 and 1992. These standards served as starting points for many states who wished to create their own.

A rather large and extended effort toward national standards was the several-faceted America 2000 program, which engaged communities through citizen-run committees. In the House and Senate, this program also provoked great discussion.

Different ideas emerged. Equity concerns arose, and what came to be called the “opportunity to learn standards,” as well as the issues of “choice” and of private-school vouchers, became sources of argument. As John Jennings summarized: “... the contention surrounding that issue [opportunity to learn standards] was to delay the legislation so long and cause so much ill will that it ultimately killed the bill.”<sup>1</sup>

Meanwhile, Secretary of Education Lamar Alexander — with the approval of Congress — urged the bipartisan National Council on Education Standards and Tests to “advise on the desirability and feasibility of national standards and tests.”

The council's 1992 report recommended national content standards and assessments based on these standards, but did not advocate a national test. The report set out criteria for creating content standards, recommending that the standards be national but not federal, and stating that they were not to become part of a “national curriculum.”

---

<sup>1</sup> John F. Jennings, *Why National Standards and Tests?*, Sage Publications, 1998.



But the report also introduced the idea of “national delivery standards,” something that received a lot of attention and was the focus of great disagreement. The premise was this: If students were to be held accountable for meeting the standards as measured by the tests, assurance was needed that instruction would give students the “opportunity to learn” the content.

The Clinton administration’s Goals 2000 legislation was the next foray into national standards. In 1994, the National Education Standards and Assessment Council was authorized, as recommended in the National Council on Education Standards and Tests report, but never established. Again, much debate and controversy ensued.

The Clinton administration proposed creation of voluntary national tests in fourth-grade reading and eighth-grade math. The National Assessment Governing Board (NAGB), which oversees NAEP, was to manage execution of the tests. Much work was done, including developing a “framework” to guide test construction and constructing actual test items. Hearings on the issue were held around the country, but the effort eventually dissolved due to lack of funding.

Also in 1994, amendments to the Elementary and Secondary Education Act (ESEA) of 1965 required states to establish content standards, tests to measure student achievement in these standards, and performance standards. Standards were still popular, and federal involvement was acceptable — but only as far as saying what had to happen in the states, rather than in some national or federal entity.

The nation’s governors and business leaders created the nonprofit organization Achieve to work with states in carrying out a standards-based reform agenda. Achieve has become an important organization, both in helping states improve their standards and in gaining collaboration around shared objectives. Given its agenda and experience, Achieve is likely to play a significant role in any further movement toward common national standards.

In 2001, through the NCLB Act, the federal government took previous efforts a giant step further. It began using the cut-points on state achievement tests, which represented attainment of “proficiency,” as the cornerstone of a test-based accountability and sanctions system for states. At this writing, the country is on the precipice of a turning point with the NCLB-reauthorization question looming over Congress and the Obama administration.



# The Current Conversation About National Standards

Although the nation has been up the standards hill and down the other side over the past two decades, some strong voices urge still another trip up the hill. Movement toward collaborating on standards setting is gaining traction, and some thoughtful efforts have taken place to examine the problems and possibilities.

These are briefly referenced below and will be mentioned again later:

- Some well-known leaders — including former North Carolina Governor James Hunt, former IBM board chairman Louis V. Gerstner Jr., former Colorado Governor Roy Romer, and others — have long been strong advocates, arguing that the nation is in crisis without standards.
- Former Secretaries of Education William J. Bennett and Rod Paige have argued that national standards and a national test are needed.
- In 2006, Chester Finn Jr., Liam Julian, and Michael Petrilli issued *To Dream the Impossible Dream*, published by the Thomas B. Fordham Foundation, with a dozen well-regarded contributors. This thoughtful effort laid out “Four Approaches to National Standards and Tests for America’s Schools” and provided a range of levels of developments and national interventions (more detail is provided later). In fall 2008, the Fordham Foundation held a debate on national standards.
- Education Sector sponsored “A Debate on National Education Standards” in 2006. The proceedings are summarized in “Five Experts Square Off,” *ES Review*, fall 2006.
- In 2007, the Nelson A. Rockefeller Institute of Government held a conference that produced the transcript published in *Intergovernmental Approaches for Strengthening K-12 Accountability Systems*.
- A July 2008 edition of *Education Daily* states: “A group of big city school superintendents addressed Congress on Thursday to retain accountability provisions in a reauthorization of NCLB, and counseled a move toward growth models and national standards. Arne Duncan, CEO of Chicago Public Schools, urged legislators to ‘set a uniform achievement standard but allow states flexibility in meeting it.’”<sup>2</sup>
- In 2007, U.S. Secretary of Education Margaret Spellings argued that “the debate over national standards would become an exercise in lowest-common-denominator politics” and would not necessarily improve the content of schooling.<sup>3</sup>
- In 2007, the National Conference of State Legislatures voted to reject the idea of common standards, saying, “We need rigorous state standards that are anchored in real-world demands.”<sup>4</sup>
- In October 2008, the James B. Hunt Institute for Educational Leadership and Policy set out a “Blueprint for Education Leadership.”

---

<sup>2</sup> Education Daily, July 18, 2008.

<sup>3</sup> Education Week, June 10, 2008.

<sup>4</sup> Education Week, June 10, 2008.

- The National Research Council of the National Academies held a series of workshops to analyze prospects, procedures, alternatives, and problems related to national standards. The most in-depth effort to grapple with both overall objectives and details of the problems, the workshops produced the 2008 summary report *Common Standards for K-12 Education? Considering the Evidence*.<sup>5</sup>
- By March 2009, the National Governors Association, the National Association of Secondary School Principals, the Council of Great City Schools, and the American Federation of Teachers (AFT) all publicly supported national standards.<sup>6</sup>

Although many statements seem to support taking action, most are sparse on the details of what to do and how to do it, as backed by solid analysis. However, the above-mentioned documents provide excellent starting points for the discussion on national action related to standards.

---

<sup>5</sup> National Research Council, *Common Standards for K-12 Education?: Considering the Evidence*, Washington, D.C.: National Academies Press, 2008.

<sup>6</sup> David Hoff, *Education Week*, March 4, 2009.

# What Does Setting National Standards Mean?

Concepts for imposing national education standards range from a general proposition that all schools help students achieve some common degree of learning, to a more sophisticated proposal that “content standards” be established and evaluated in periodic surveys, such as those carried out by the AFT and by the Fordham Foundation.

The rigor and quality of existing state standards vary greatly, and states have different mindsets about what content standards are intended to do. One state may see standards as an expression of high aspirations for how much students should know. Another state may view standards as a way to make realistic judgments about what is possible for students to know, given that state’s experience with its schools. Still another state, seeing tests designed to measure school effectiveness and to trigger sanctions, may set modest standards to ensure perceived success. These and other differences in thought and approach surely contribute to the wide differences revealed in evaluations of the states.

Once a state determines what the standards should do, it must decide how demanding to make them. Before doing this, those creating standards must understand how wide the distribution of achievement is in that state’s schools in any one grade. For example, where the spread of achievement is wider — perhaps as much as three grade levels from the bottom tier to the top — setting a single performance standard becomes problematic.

Some serious advocates of national standards have thought principally in terms of a more rigorous curriculum. This mirrors the model advocated by Diane Ravitch and the NCTM in the late 1980s.

Many advocates of national standards want such content standards, and a standardized test based on those standards. Others prefer to jump to the proposition that a good national test is needed, period. The most general belief is that a “good” set of content standards can be fashioned and that it is possible to determine the level of student performance from standardized testing. This may be achievable, but given the nation’s unwillingness to invest in high-quality assessments that go beyond filling in the bubbles, this idea faces many hurdles.

Lauren Resnick, a leading authority on standards and assessments, summed it up this way: “The tests are not aligned to their own state standards in all but a very few cases. . . . Most of the state tests do not test the high level, intellectual demands that we were after when we set up the standards.”<sup>7</sup>

Those who focus primarily on a standardized test do so because it provides test scores for a sanctions-based accountability system in a way that simply describing course content does not. For example, in the aforementioned Rockefeller Institute of Government conference, the morning agenda was “Models for Setting Academic Standards” and the afternoon agenda was “Approaches to Testing Oversight and Accountability” — yet, most of the morning session addressed testing, and the afternoon was all about testing.

Another key question is whether national standards should be voluntary. Should they serve merely as useful models or as mandatory principles, enforced through the Elementary and Secondary Education Act or other legislation? Before the country can seriously consider national standards, it must seriously address these difficult questions around the topic.

---

<sup>7</sup> Lauren Resnick, “Five Experts Square Off,” *Education Sector Debates*, 2006.

## Some Recent Developments

The education system is largely a public organism operating in the public sector: created, funded, controlled, and administered by a local government, under a state constitution and state laws, with some financial support and requirements under federal law. Governmental and law agencies impose existing standards by defining a curriculum's content, administering tests to see if students know the content, and applying incentives or sanctions accordingly.

Before NCLB, many states provided both incentives and sections based on tests. When NCLB was passed, the new law simply used the cut-points designated as representing "proficiency" on these tests as a basis for applying sanctions. However, after the law passed, states were free to change their cut-points.

Much of the discontent with NCLB has been over the large variation in these state-set cut-points. Although many states have set what are considered low cut-points, former Secretary of Education Spellings has stated that creating a nationwide cut-point will only perpetuate the problem, likely establishing a single standard that is too low.

Considerable discussion has taken place around how to establish collaboration outside the federal government to set education standards, how to shape that effort, and how ambitious to make it. In 2006, the Thomas B. Fordham Foundation published *To Dream the Impossible Dream: Four Approaches to National Standards and Tests for America's Schools*. The 44-page work included contributions from a dozen well-known people in the field, including Chester E. Finn Jr., Liam Julian, and Michael J. Petrilli. A well-attended forum followed in Washington, D.C.

The four approaches offered and discussed are as follows:

1. **The Whole Enchilada.** The federal government imposes standards on the states, and standards are used for accountability. The Fordham Foundation concluded that this approach was not politically feasible. On the merits of this approach, forum participants had varied views.
2. **If You Build It, They Will Come.** This is a voluntary version of the first model, with the federal government or a private group developing standards and providing incentives. Contributors judged this approach "maybe" politically feasible.
3. **Let's All Hold Hands.** Through this approach, states are encouraged to join one another in developing common standards, possibly with the federal government providing incentives, but no more. Contributors also judged this approach "maybe" politically feasible.
4. **Sunshine and Shame.** Called the "least ambitious model," this approach ensures that state standards and tests are "more transparent" by making them easier to compare with one another and with NAEP. Contributors considered this approach feasible.

Some of the same people gathered in Chicago in 2007 with Richard Nathan and the Nelson A. Rockefeller Institute of Government to consider the aforementioned "Intergovernmental Approaches for Strengthening K-12 Accountability Systems." The edited and published transcript includes a framework paper written by Allison Armour-Garb of the Rockefeller Institute and distributed to conference participants.

Discussions among the 37 participants were broad and far ranging, but informed more than they concluded. The Rockefeller Institute promised to use the proceedings for a further effort and to produce a new paper.

Any effort to move forward on national standards will benefit greatly from recent work of the National Research Council (NRC) of the National Academies. The United States has no more respected source of scientific knowledge and research than the NRC, which conducted a series of workshops and summarized the results in its 2008 publication *Common Standards for K-12 Education?: Considering the Evidence*.

The following is a list of topics covered in the NRC publication, which will be revisited later in this report:

- the degree of variability in content and performance standards and in “paradoxes,” including views from five states
- a description and consideration of available options
- analysis of the quality of content standards and of their impact on teaching and learning
- estimated costs
- political and legal considerations
- perspectives of researchers and elected officials on implementation of common standards

It is hard to read the NRC report carefully and not come away with an understanding of the important questions, needs, and considerations of feasibility and desirability around standards.

In a November 2008 paper titled *The Role of Assessment in Federal Education Programs*, commissioned by the Center on Education Policy, W. James Popham provides a “serviceable framework for rethinking an appropriate federal role in U.S. educational testing.”

Popham addresses what role the federal government should have, what the measurement missions of tests should be, and five gradients of federal involvement, ranging from zero to total federal control. With an academic career in educational measurement, Popham for years has been analyzing and regularly writing and publishing on what has been happening during the standards-based reform and test-based accountability period. His paper includes a review of the federal role over the past 50 years.

The organization Achieve has made the longest collaborative effort, with the widest reach. Achieve is the result of a joint endeavor among the nation’s governors, chief state school officers, and CEOs of large corporations, informed by earlier efforts that created the National Education Goals Panel.

Another base of experience is the New England Common Assessment Program (NECAP), through which New Hampshire, Rhode Island, and Vermont are implementing a common standards-and-assessment program. A principal incentive for these three small Northeast states is the considerable amount of money they can save through collaboration. While the cost of having separate standards has not yet motivated the remaining states, the recession of 2008 – 2009 may help push more of them to work together.

Achieve, the National Governors Association, and the Council of Chief State School Officers recently launched an International Benchmarking project that could help states set their own standards or attempt to set common standards. To be truly useful, the work must specify whether it is concentrating on setting standards for what should be *taught* or for how much should be *learned*.

Additionally, Achieve, the Data Quality Campaign, the Education Council, and Jobs for the Future have created a partnership to support the work of states through the College and Career-Ready Policy Institute, supported by the Gates Foundation. Although this partnership focuses on high school, it may prove useful as a starting point. By first setting standards at the high school level and then tying them to real-world outcomes, groups can work back to the elementary grades to create avenues of progression.

In January 2007, Senator Christopher Dodd and Representative Vernon Ehler introduced a substantial piece of legislation in a bipartisan bill. The goal was to provide incentives for

states to adopt math and science content standards developed by the governing board of NAEP. The bill was endorsed by about 40 organizations. That same month, Senator Edward Kennedy introduced a bill to encourage states to benchmark their standards and assessments to NAEP.

Although this roundup by no means captures all of the advocacy, analysis, and proposed approaches, it is highly representative. The pot is filling and beginning to simmer as the new Obama administration settles into office.

# Recognizing the Wide Variation in the U.S. Education System

The wide variation in the U.S. education system is why many advocate a “common” standard for what is taught, when it is taught, and how to test student performance.

Variation is a staple of the U.S. public education system, given how the country created and expanded its system. Its start was not at the state or national level, but at the local level, with the establishment of the first school system in Massachusetts during the Colonial period. The U.S. Constitution does not even mention the provision of education, nor do the Federalist Papers.

Forces developed over time that encouraged commonality, and more commonality has developed than is typically recognized. In his history of the Colonial period, *The Americans: The Colonial Experience*, Daniel J. Boorstein commented on how Americans, despite their different dialects, developed enough commonality in language and pronunciation to communicate with one another. This contrasted with England, he said, where citizens with distance between them could barely understand one another, if they could at all. Boorstein attributed this difference to American “schoolmarms,” who taught a standard form of English.

In the early American schools, all students likely studied from McGuffey Readers. In fact, historically, U.S. textbooks have included similar content, competing instead by varying their presentations and incorporating emerging trends in educational preferences.

Moreover, a few common tests with names long in existence, such as the Iowa Test of Basic Skills, were used throughout the country and shared the same test questions. Student performance varied greatly in any one grade, with these “norm-referenced” tests highlighting exactly how

much student achievement varied. Students and parents knew whether a ninth-grade student was reading at the seventh-grade level or the 11th-grade level, and “grade-level achievement” simply referred to average student scores in a particular grade.

But there were large elements of variation within the uniformity. Mixed up in this was a distinction, still not often emphasized, between the content of class instruction and the achievement levels attained by the students in the classes. These are very much related but are not the same. This variation increasingly has been the focus as the nation has fretted over achievement outcomes, particularly as concern grew about the wide and persistent achievement gaps among racial/ethnic and socioeconomic groups and the United States’ declining educational advantage in a competitive global economy.

To encapsulate the extent of variation in U.S. classrooms, another distinction must be more carefully addressed: having high standards — i.e., creating a rigorous curriculum — vs. standardizing what goes on in classrooms throughout the country. The case for raising achievement expectations is different than for increasing uniformity or standardizing the content and delivery of instruction across the nation.

## Variation in Content

The AFT and the Thomas B. Fordham Foundation have studied content standards and tests in all states. They report wide differences in such areas as quality of standards, assessments to measure student performance, alignment between content standards and tests, and — to some extent — alignment of content standards and tests with the actual “delivered curriculum.”



Lauren Resnick, in collaboration with others, also has gone into considerable depth about how states implement standards. In an Education Sector Debate in late 2006, she summed up some of what she learned:

*There can be differences of opinion about what good standards are, but no one will disagree that there is variability in quality. That means that we have very different state standards with the same consequences attached to them. ... The tests are not aligned to their own state standards in all but very few cases. ... Most of the state tests do not measure the high-level, intellectual demands that we were after when we set up the standards ...*<sup>8</sup>

Despite Resnick's conclusions, and regardless of whether flawed systems produce quality data, federal law continues to sanction schools uniformly.

The National Research Council report discussed previously sums up what has been learned about variability in content and performance standards. Using research on 14 states in English/language arts/reading from Andrew Porter and his colleagues, the NRC measured variation via an index of 0 to 100, with 100 representing perfect alignment.<sup>9</sup>

The range of variation was huge: from less than 1 between Maine and Wisconsin for grade eight, to highs at just under 50 between Ohio and California. Overall, the researchers “found little evidence to support the hypothesis that there is a de facto national curriculum.”

Porter and colleagues also looked at variation from grade to grade in the degree of overlap in

standards for each grade. They found that “the alignment of topic coverage within states from grade to grade ... is generally greater than the degree of alignment across states in the material they cover at particular grades.” They did, however, find indications that a few core areas were consistently covered among the states — “a small de facto common core curriculum.”

Some in the national standards movement have advocated aiming for creating some common core in the curriculum, without trying for total uniformity. It would be useful to see if what exists now could be identified and described from the research by Porter and colleagues. But expanding standards beyond just a common core also has proponents.

The NRC report includes discussion of the work of Barbara Reys, who observes that “it’s really the decisions about what you want to focus on in the individual grades that are the tough ones.” Even when states include the same content, that content may be taught in different grades, she finds. This could result in different progressions from grade to grade.

Resnick looks beyond the whole orientation toward individual grades and specification of content to be covered in each grade. Basing her conclusions on 15 years of research in testing and psychometrics, she says: “The new idea is to set a graduated set of goals for instruction that describe a *learning progression*: an ordered sequence of goals that a student would be expected to meet if he or she was successful in a well-conducted instructional program. There might be three to five major goals each year.”<sup>10</sup>

---

<sup>8</sup> Resnick, 2006.

<sup>9</sup> Andrew Porter et al., “Is There a De Facto National Curriculum? Evidence from State Content Standards,” Referenced in National Research Council, 2008, p. 23.

<sup>10</sup> Lauren Resnick, et al., “Standards-Based Reform: A Powerful Idea Unmoored,” in *Improving on No Child Left Behind*, Richard Kahlenberg, editor, Century Foundation Press, 2008.

Assessment, she continues, would occur after instruction is completed on a given goal. Presumably, goals could overlap grade levels. Resnick points out, however, that only a few such learning progressions have been identified and documented. She also notes that the concept of progression from grade to grade is not a settled matter: The United States, Resnick says, teaches ordinary fractions first and decimals a year or two later, while teachers in France and other European countries do the opposite. She adds: “These different teaching sequences will entail different testing sequences.”<sup>11</sup>

The role of textbooks in education is important to any discussion about common standards. A teacher does not teach from a volume of standards provided by the state in grade eight mathematics. Students read from textbooks and use workbooks that go along with them, to a greater or lesser degree.

Although states and school districts select which texts will be used, they do not directly decide what is in these books. Teachers do use supplementary materials, of course. And while there are 47 sets of state standards for each grade and subject matter (three states have common standards), there are not 47 sets of textbooks customized to the standards of each state. This matter is rarely addressed in standards-based reform formulations or in discussions of common standards, but seems to warrant greater attention.

Given the large reliance on grade-specific textbooks, and given that only a few are available from different companies for any one subject and grade, and that publishers compete on attractiveness and presentation more than on content, it seems reasonable to suppose that publishers, intentionally or not, helped establish

content standardization even before states systematically defined it. In any event, textbooks must be recognized as principal mediators between what is in state content standards and what is taught in classrooms.

Reys also discusses the relationship of different content placement in grades and challenges faced by textbook producers. These differences may lead publishers toward accommodating variability in content rather than seeking commonality. Reys observes that “only 4 of 108 possible learning expectations for fourth graders were common across 10 states — suggesting that a textbook publisher might choose to incorporate all 108 of them. Since the content of textbooks has a significant effect on teachers’ instructional plans, this lack of overlap becomes a self-reinforcing pressure against curricular focus.”<sup>12</sup>

A set of content standards and objectives that excludes the nature and dynamics of the textbook market ignores a key force in the issue of national standards.

### Variation in Performance Standards

Different terms are used to describe the standards for tests that measure performance, but the most operational is the test score cut-point labeled “Proficient.” Much discontent that has developed around NCLB involves the considerable variability among states of cut-points defined as “Proficient.” NCLB aims for all states’ students to reach a state-set cut-point by 2014 and to meet a trajectory set by each state every year leading up to that goal.

Another point of contention is that the states have widely varying trajectories, with some moving in a conventional straight line and others moving more gradually in early years and then

---

<sup>11</sup> Resnick, 2008.

<sup>12</sup> Barbara Reys, cited in National Research Council, 2008.

faster in later years. The latter group is thought by some to be buying time in hopes that the law will change. Chester Finn has referred to these states as having “taken out a balloon mortgage.”

NCLB’s principal focus is, simply, variations in achievement expectations among states, represented by variations in cut-points at any one time. However, it is hard to compare one state with another since each has its own tests and reports results on its own achievement scales.

The Fordham Foundation and the Northwest Evaluation Association (NWEA), which conducts testing in many states, set out to compare states by level of difficulty of their proficiency cut-points. The study drew on 830,000 students in 26 states where students had taken both the state test and the NWEA test, which provides a basis for comparing students on both tests.<sup>13</sup>

The researchers found “enormous” variability among participating states. At the low end of cut-points, 94 percent of students passed; at the top of the range, only 23 percent passed.

After NCLB was passed, many said states engaged in a “race to the bottom” to lower cut-points and avoid sanctions to their schools. But Michael Petrilli of the Fordham Foundation said there was a “walk to the middle,” based on findings from the above study, which looked at scores and cut-points over two points in time. This resulted either from the states changing cut-points or from changes in achievement distributions.<sup>14</sup>

Given that all states now participate in NAEP reading and mathematics assessments, variability

in student performance can be seen by looking at scores on the NAEP scale — for example, the huge difference in scores between students at the 10th and 90th percentiles of the same grade.

Some see the variation as reason to have a common cut-point, and view the fact that so few reach anywhere near the NAEP “Proficient” cut-point as an argument for raising standards. Others see it as the result of setting the NAEP cut-point unrealistically high.

NCES contracted with Henry Braun, then at Educational Testing Service (ETS), to learn more about how NAEP provides insight on state standards. For 34 states, Braun “mapped” the state test’s cut-points for proficiency onto the NAEP scale.

NCES has used results of the mapping process to report where each state’s proficiency cut-point falls on its own test, compared with where it falls on the NAEP scale for the 2005 state reading and mathematics assessments for grades four and eight. The overall conclusion: “For each of the four subject and grade combinations, the proficiency standards vary widely, spanning a range of 60 to 80 NAEP score points. Although there is an essential ambiguity in any attempt to place state standards on a common scale, the ranking of the NAEP score equivalents to the states’ proficiency standards offer an indicator of the relative stringency of those standards.”<sup>15</sup>

The above statement implies that uniform “stringency” among states is desirable and that, therefore, differences among states’ economies and resources should not affect the cut-points they

---

<sup>13</sup> Drawn from the summary provided in National Research Council, 2008, pp. 15-19.

<sup>14</sup> In *The Accountability Illusion*, 2009, the Fordham Foundation took the comparisons to the individual school level, and selected from around the nation 36 schools that varied by size, achievement, and diversity. Fordham found that similar schools varied on whether they met the Adequate Yearly Progress requirement because they were in different states, and different states have different performance standards.

<sup>15</sup> National Center for Education Statistics. *Mapping 2005 State Proficiency Standards onto the NAEP Scales* (NCES 2007–482). U.S. Department of Education, National Center for Education Statistics, Washington, D.C.: U.S. Government Printing Office.

set on a common national scale. This proposition may be basic to the argument for establishing national standards, and it will be discussed later in this report. Another way of looking at variation in state cut-points is by considering the different resources and capabilities states have for their education systems — as might happen when comparing highly industrialized nations and less-developed ones.

So have the states set achievement targets that reflect such differences? Clearly, the answer is no. Table 1 shows the percentage of students who are at or above the state-set cut-point as mapped onto the NAEP scale for the NAEP assessment in each state, rather than just for the NAEP scores for the country as a whole, as are shown in the last two columns of Table 1. The numbers vary widely but do not seem to be affected by the strength of the state economies.

One striking contrast is between North Carolina and South Carolina, states located next to each other and with practically the same average NAEP scores. But North Carolina's cut-point is set low, with 88 percent reaching or exceeding it, and South Carolina's is set high, with just 30 percent reaching or exceeding it.

The comparison addressed only a “performance standard,” in terms of the cut-point on the test, not a standard for what should be taught. As pointed out earlier, both states have nearly the same average NAEP scores.

Few states come anywhere near reaching the NAEP “Proficient” level, which has received criticism as having been set too high. But many states reach or are near the NAEP “Basic” level, as can be seen in Table 1.

Table 1 shows how the percentage of students reaching state cut-points compares with the percentage reaching the NAEP achievement levels of “Basic” and “Proficient,” and also shows the great variation among states. Variation abounds in terms of performance standards.

Each of these state-set proficiency cut-points is set by well-known procedures, of which there are several. Many people and educational measurement experts are involved in setting state proficiency cut-points, with the U.S. Department of Education reviewing the states' processes. The data in Table 1 suggests that it is extremely difficult to establish standards and objective processes for deciding what constitutes acceptable performance. Further, NAEP has long been criticized in the scientific community for how it sets its proficiency cut-points. Much subjective judgment obviously is involved.

As far as this author has seen, Table 1 provides states with their first look at how state cut-points compare with the NAEP scale for their state. This information could form the basis for a conversation on how states vary and on how to view states' scores within a standards framework.

North Carolina seems to have established low expectations in how much it is trying to raise the state cut-point, while South Carolina appears hugely ambitious in its efforts. One approach toward reconciling the two might be to choose a percentile within the states and then ask all states to aim for that level — for example, the 25th percentile, which is about where Connecticut and Iowa are. This type of standard is akin to expecting states to make the same *relative* progress based on their own distribution of scores. But of course, it leaves open the question of what percentile to choose as a standard.

**Table 1**  
**State-set Proficient Standards Compared With NAEP Achievement Levels**  
**Grade 8 Reading, 2005**

| State                | Percentage of Students At or Above State-set Standard for Proficient | Percentage of Students At or Above Basic on NAEP | Percentage of Students At or Above Proficient on NAEP |
|----------------------|--|--|---|
| Alaska               | 81   | 70   | 26  |
| Arizona              | 64   | 65   | 23  |
| Arkansas             | 57   | 69   | 26  |
| California           | 39   | 60   | 21  |
| Colorado             | 86   | 75   | 32  |
| Connecticut          | 75   | 74   | 34  |
| Delaware             | 44   | 80   | 30  |
| District of Columbia | 81   | 45   | 12  |
| Florida              | 43   | 66   | 25  |
| Georgia              | 83   | 67   | 25  |
| Hawaii               | 37   | 58   | 18  |
| Idaho                | 82   | 76   | 32  |
| Illinois             | 63   | 75   | 31  |
| Indiana              | 67   | 73   | 28  |
| Iowa                 | 73   | 79   | 34  |
| Kansas               | 78   | 78   | 35  |
| Louisiana            | 54   | 64   | 20  |
| Maryland             | 67   | 69   | 30  |
| Mississippi          | 56   | 60   | 18  |
| New Jersey           | 74   | 80   | 38  |
| New Mexico           | 52   | 62   | 19  |
| New York             | 49   | 75   | 33  |
| North Carolina       | 88   | 69   | 27  |
| North Dakota         | 72   | 83   | 37  |
| Ohio                 | 79   | 78   | 36  |
| Oklahoma             | 71   | 72   | 25  |
| Oregon               | 64   | 74   | 33  |
| Pennsylvania         | 64   | 77   | 36  |
| South Carolina       | 30   | 67   | 25  |
| Tennessee            | 87   | 71   | 26  |
| Texas                | 83   | 69   | 26  |
| West Virginia        | 80   | 67   | 22  |
| Wisconsin            | 86   | 77   | 35  |
| Wyoming              | 40   | 81   | 36  |

Source: The first column numbers were calculated at ETS based on the state cut-point as established in "Mapping 2005 State Proficiency Standards Onto the NAEP Scale" (NCES 2007-482), June 2007. The numbers in the second and third columns are from NAEP.

Also worth noting: If all students are moving up, those both above and below the 25th percentile will be among them. This has not been well recognized in the current accountability approaches, and gaps between subgroups have stayed relatively constant even as more students reach the required state proficiency cut-points. Certain policies and programs have actually *increased* inequality at the high school level, intentionally or not. Examples are the large number of Advanced Placement® courses and the dual enrollments in high school and early college programs.

The currently favored high school reform approach of increasing rigor so that more students are prepared for college and pass college-placement examinations likely will raise achievement to higher levels, particularly for the already more advanced students who are best prepared to take advantage of these new opportunities. Efforts are also beginning to push higher standards on middle schools so students there will be better qualified for tough high school courses.

This author’s earlier point about state *performance standards*, or test cut-points, having no proven relationship with differences in the states’ economic and social capital should not be confused with the states’ wide variations in *student achievement*. Research has well established that variations in average scores among states are highly correlated with indices of variation in such state resources.

### Variation in Student Achievement

The United States may very well deal with more inequality of educational achievement than any other developed country. This is one reason states feel pressured to establish a level of achievement that all students must reach, yet challenged by how to do so.

The most comprehensive view of variability in the nation as a whole is in Figure 1 on the following page. NAEP’s “long-term trend” series provides achievement scores for students ages 9, 13, and 17.

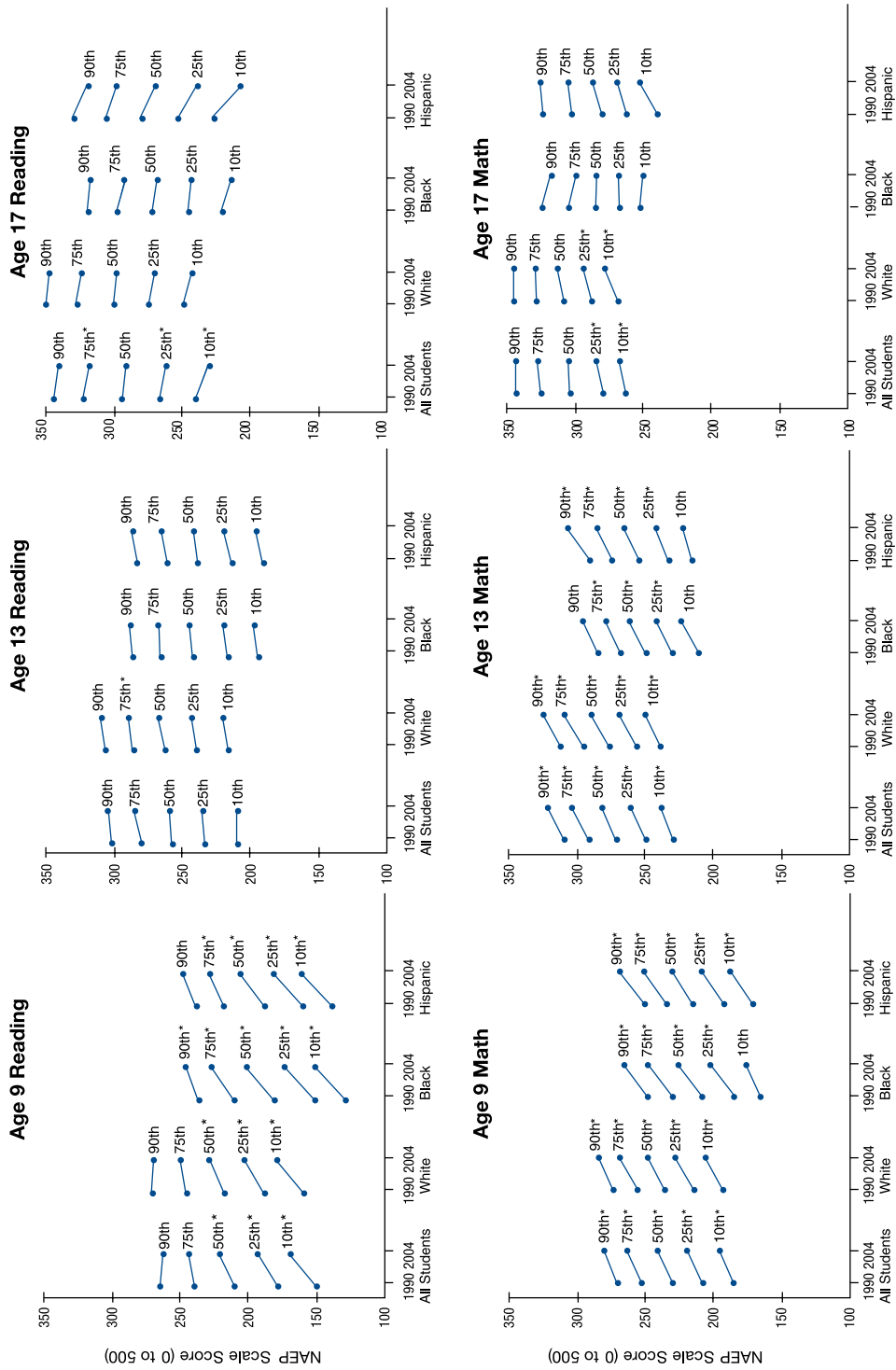
The chart offers a glimpse into score distributions for all three age groups, as well as a way of viewing score trends over time, pinpointing where change is and is not occurring. It also separates out score distributions by race and ethnic group. Ultimately, Figure 1 shows the width of the distribution at any one age and the amount of overlap among students in the three age groups.

After reviewing this chart, one quickly realizes that finding one point to set as a standard is a little like playing Pin the Tail on the Donkey — haphazard.

The wide range of scores is readily apparent. So is the gap among racial/ethnic subgroups. But perhaps the most striking discovery of all is the overlap in achievement among different student age groups — four to nine years apart in the number of grades in the school system. By placing a ruler straight across the chart, readers will see that the bottom fourth of 17-year-olds only do as well as the top tenth of 9-year-olds. Perhaps the term “national educational improvement goals” is more accurate than “national standards.”

Although the chart clearly shows a wide spread in achievement scores, how does one grasp its magnitude? How does one compare the size of the spread from one group to another, or see how the spread is changing over time? One simple way is to look at the difference in scores of students at the 10th and 90th percentiles.

**Figure 1:** Percentile Distribution of NAEP Reading and Mathematics Scores, by Age and Racial/Ethnic Group, 1990 and 2004



\* Indicates statistically significant difference from 1990 to 2004.  
Source: NAEP special tabulations prepared by ETS.



Table 2 shows such differences in reading scores for 2005 (the year chosen because the 2007 assessment did not include 12th graders).

**Table 2**

**Difference in Scale Scores Between 10th and 90th Percentiles, NAEP Reading, 2005**

| Students       | Grade 4 | Grade 8 | Grade 12 |
|----------------|---------|---------|----------|
| All            | 92      | 76      | 98       |
| White          | 83      | 64      | 94       |
| Black          | 87      | 67      | 92       |
| Hispanic       | 91      | 70      | 91       |
| Asian American | 91      | 71      | 98       |

Although there is much similarity in spreads at the fourth and 12th grades, the spread is considerably smaller in the eighth grade. The score spreads among the subgroups are near the same magnitude (a test of statistical significance was not run); the range is approximately 70 to 90 points.

Here is some additional background: The average reading scores for all students in 2005 was 219 for grade 4, 262 for grade 8, and 286 for grade 12. So the difference in average scores between grades 4 and 8 was 43 scale points, and between grades 4 and 12 was 67 points.

The spread in scores is much larger in any one grade than between the averages in grades 4 and 8. *In fact, the spread in scores within grade 8 is as large as the difference in average scores between grades 4 and 12. And the spread is considerably greater within grades 4 and 12 than the difference in average scores between these grades.*

Another way to view the wide variation in what students know and can do is by examining an item map that shows the kinds of mathematics tasks involved in the 2000 NAEP mathematics assessment for grades 4 and 8, the NAEP proficiency levels, and the points along the score scale where selected groups of students score. A few highlights of this comparison, shown in Figure 2, include:

- The average score of eighth graders in the District of Columbia is about even with the average score of White fourth graders nationally
- The top-scoring state, Minnesota, scores, on average, as well as the highest scoring racial/ethnic group – Asian/Pacific Islanders
- American Indian, Hispanic, and Black eighth graders score below the “Basic” level, on average.<sup>16</sup>

These kinds of data and analyses make contemplating a common, national standard a daunting proposition.

The struggle to address this wide spread in achievement affected the National Educational Goals in 1989, when the president and state governors set a national goal of raising achievement in all four quartiles and narrowing gaps by race and ethnicity. The National Education Goals Panel tracked progress accordingly.<sup>17</sup>

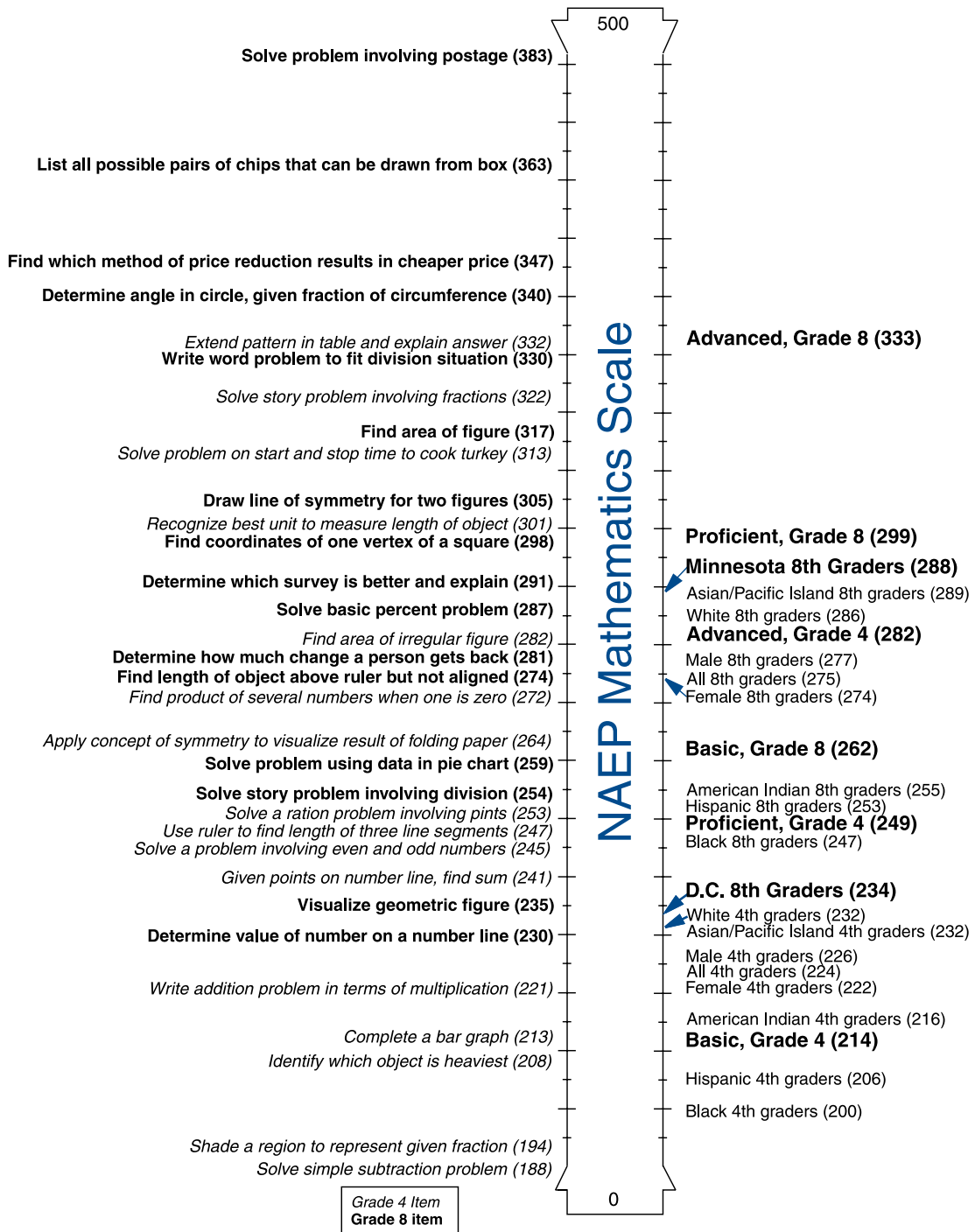
A parallel exists in what former AFL-CIO president George Meany said about the wage goals of organized labor many decades ago. “More,” he stated simply.

Addressing the huge variation in achievement is one of several critical factors in the success of any effort to set national standards.

<sup>16</sup> Coley, 2003.

<sup>17</sup> This author wrote two reports that the goals panel published to track progress in these terms. An updating of this work can be found in *Windows on Educational Achievement and Inequality*, by Paul E. Barton and Richard J. Coley, ETS Policy Information Center, 2008, pp. 29-30.

**Figure 2:** Map of Selected Items on the NAEP Mathematics Scale and Average Scale Scores for Selected Groups of Students, 2000



Source: James S. Braswell et al., *The Nation's Report Card: Mathematics 2000*, Washington, D.C.: National Center for Education Statistics, August 2001.  
 Note: The position of an item on the scale represents the scale score attained by students who had a 65 percent probability of successfully answering the item. (The probability was 74 percent for 4-option questions and 72 percent for 5-option questions.)

# The Importance of Widespread Variation in Achievement to the National Standards Discussion

The focus of the national standards debate has been on differences in the content of instruction and curriculum across the nation, and the degree to which students have mastered that content. Data this author has seen shows no clear way to identify variation in classroom-delivered content. Little is known beyond the course and subject titles, except from a few in-depth studies by researchers seeking to determine how closely “delivered” curricula line up with states’ content standards and/or tests.

Variation in the content of instruction is important to know because concern about perceived undesirable variation drives the demand for national standards. Conclusions about that variation must be imputed from a welter of data that encompasses other types of variation; it also can be gathered from empirical knowledge of variation dictated by the necessity to adapt content in any one subject and grade to fit the degree of prerequisite knowledge of students at some starting point — eighth-grade math, for example.

How much must a teacher backtrack to bring entering students up to what might be considered the place to start? How much can a teacher jump ahead because of the high knowledge base of entering students? A brief review of this variation follows.

Students enter first grade with unequal levels of cognitive development and achievement. Despite what is considered a standard first-grade instructional program in a state or district, teachers in any given school must deal with whatever reality they confront; thus, content varies from class to class and school to school.

School districts encompass a variety of neighborhoods where conditions can vary greatly. Students may include children born to mothers who are high-school dropouts and have no resident male partner. Such children may have been underweight at birth, receive poor nutrition, and experience little verbal interaction. Students also may include children born into professional families with two parents. These children may have received regular, nutritious meals, participated in much verbal interaction with parents, and been nurtured in quality preschools.

Children from both types of families will begin school with different knowledge bases and skill sets. They all will advance in knowledge and skills, because all children can and do learn. Attempts at accelerating the growth for those who start behind and stay there may be limited by the amount of time in a school day and year, by resources too limited to expand instructional time, and by the generally meager results so far achieved in supplementary services. Such students will play catch-up to students ready to sprint ahead.

Variability in the state-prescribed content of instruction also seems to affect differential achievement, even after considering class and school adjustments for entering students’ different achievement levels. The question is, how much of the difference relates to the state-prescribed content and how much relates to other powerful forces that influence the content of instruction?

Variation in student achievement also arises from observed differences in school experiences. The instructional content may be identical and may meet a set of given standards. But the teacher may be inexperienced or uncertified, may not have a major or minor in the subject matter being

taught, or may be one of several teachers in that subject the student sees that year, due to high turnover. Furthermore, students may be in a school where a considerable gang presence, fear, physical fighting, or much classroom disruption exists.

Differences in summer activities present another obstacle to having a common rigorous curriculum for all students. Students returning from the summer have had widely varying experiences, ranging from reading, attending camp, and traveling, to primarily watching TV and staying at home. These experiences can affect whether a student enters the school year in a refreshed or regressed state.

Seekers of common standards tend to concentrate on the elements of high standards for instruction content and express concern over variation in this content. Content may be unchallenging, and evaluations conducted

by the AFT, the Fordham Foundation, Achieve, and independent researchers show that content standards created by the states are sometimes weak. Efforts should be made toward improvement but also must be mindful of the many forces at work that influence substandard student performance and its large variation.

If all children entered school as empty vessels, and if all schools poured into children equal amounts of knowledge, this report's discussion would be unnecessary. But children do not enter schools as empty vessels, and they have widely varying knowledge bases and developed abilities, so the content of instruction varies. This must be acknowledged.

Instruction can and should be improved, so the question is: What constructive steps can this diverse country, with its history of local and state control of education, take to resolve the continuing debate about what and how to teach?

# Who Bears Responsibility for Meeting Standards?

If students are to achieve high goals before they leave 12th grade, who should accept responsibility for ensuring that the students' goals are met? If national content standards are established and there is a national test, who will be held responsible for students reaching a performance standard on the test? The students? The schools? The states? The nation?

Most of the literature proposes that schools bear this burden, with little attention to the role that the nation's social and economic institutions play in student achievement. The national test being advocated to complement national standards is discussed as a way to see if schools are doing what they are expected to — just as state tests are used to determine compliance with NCLB.

Schools deliver formal education that is instrumental to imparting desired knowledge and skills. So to raise standards down the line, schools will be critical. But are they the only factor? Is the wide distribution of achievement scores among students in the same grade solely the result of teaching or curriculum quality? Is variation in the teaching of math the only reason that one-fourth of 17-year-olds perform no better than the top 10th of 9-year-olds in math?

Student achievement has always been framed by the results of standardized tests, which typically are based on the cumulative abilities and knowledge of students at one point in time — usually at the end of a grade. Until recent years, “norm-referenced tests” provided most comparisons of student achievement at district, state, and national levels. Some changes occurred after Robert Glaser introduced the concept of “criterion-referenced” tests in 1963.

Glaser's idea was to set goals of what and how much students needed to know, and then to test students to see how close they came to reaching these goals. This thinking formed the basis for the standards-based accountability systems that states began developing in the 1990s; it also became the basis for sanctions in NCLB.

Although NCLB required criterion-referenced testing, many states continued to use norm-referenced tests, responding only gradually to NCLB requirements. And while some state accountability systems and NCLB spelled out sanctions for ineffective schools, *few measures of student progress during the school year were in use*. Instead, measures highlighted how much, in total, students knew about a subject at the end of a grade.

In schools judged as low performing, substantial research has clearly shown low correlation between end-of-year tests of total knowledge and systems that measure how much students *learned* in school over the nine-month school year. The latter requires an entirely new system, on which a number of measurement experts have been working. Tennessee has used a “value-added” measurement system for accountability for more than a decade. This author in other work has summarized much of the research on the value-added approach.<sup>18</sup>

The highly respected NAEP measures total knowledge of a subject at one point in time, with educational progress tracked over time. Changes in NAEP scores generally are believed to reflect changes in school quality. As discussed above, NAEP's “Proficient” cut-points are frequently compared with states'; this comparison is not called the “National Assessment of Progress in

---

<sup>18</sup> Paul E. Barton, *Unfinished Business: More Measured Approaches in Standards-Based Reform*, ETS Policy Information Center, January 2005, and Paul E. Barton, “Failing” or “Succeeding” Schools: *How Can We Tell?*, American Federation of Teachers, 2006.

Achievement in School” — which perhaps could be the fitting title for a companion report to the ones now relied upon.

Many comparisons can be made from measuring knowledge growth during the school year with total knowledge. Data are available from the state of Tennessee, which has long used both approaches in its state testing systems. And NAEP data can provide striking illustrations for the nation and the states: A basic redesign of NAEP in the early 1980s resulted in the tests’ ability to track changes in achievement for a cohort of students and to report the students’ status at a point in time. An assessment is given to a national sample of fourth graders and, four years later, to a national sample of eighth graders to show gains.

In 1998 and 2003 reports, the ETS Policy Information Center looked at such growth.<sup>19</sup> Both reports found similar results, and the basic pattern today is probably the same. In general, these analyses found that despite wide differences in average scores among racial/ethnic groups, there was little difference in the “gains” made by these groups between fourth and eighth grades. When this growth among states was examined, many of the states ranking highest in average score dropped considerably down the list when “gain” scores were examined.

Regardless of how and when it is measured, learning is basic to the human experience — and it took place well before there were formal schools for children to attend. Everyone knows the story of Abraham Lincoln, for example, who had little more than a year of formal schooling as a child. Lincoln’s experience, of course, is not typical. The following description of learning in American colonies in the 17th century from Robert Bellah,

et al., in *The Good Society* provides a more general overview:

*In seventeenth-century Massachusetts, adult male literacy was about 80 percent, twice the adult rate for males in England at the time, and adult females had a literacy rate of 60 percent. Thousands of books were imported into the colonies, and many more were published here ... so we were a literate people, capable of reading and conversing about complex issues of religion and politics. Yet schools, except for a tiny minority ... were not a major part of the lives of most people ... It was the whole community that educated: the home, the church, the voluntary association and local politics had an educative function at least as important as the schools.*<sup>20</sup>

One way to differentiate the contributions made to education by schools, as opposed to individuals, other organizations, and life circumstances, is to adopt a “value-added” approach to measuring growth in achievement. This work has been left largely to the educational measurement and psychometric experts, and “black boxes” are the result. If we have some national measure of the knowledge that students gain in school, the measure needs to be transparent and useful to teachers in diagnosing student needs and providing appropriate instruction to meet those needs.

To this author’s knowledge, no focused effort has established what constitutes insufficient, acceptable, or exemplary *gains in knowledge* within a school year, or what gain would allow all subgroups to reach the same point at the end of a school year and thus close the achievement gap.

There is no official record of how much students in all demographic subgroups, or even average students, gain while in school. The

---

<sup>19</sup> Paul E. Barton and Richard J. Coley, *Growth in School: Achievement Gains from the Fourth to the Eighth Grade*, Policy Information Report, ETS Policy Information Center, 1998; and Richard J. Coley, *Growth in School Revisited: Achievement Gains from the Fourth to the Eighth Grade*, Policy Information Report, ETS Policy Information Center, 2003.

<sup>20</sup> Robert N. Bellah, et al., *The Good Society*, Vintage Books, 1991, pp. 146-147.

Northwest Evaluation Association perhaps has the best potential database for this information; it tests in many states, both at the beginning and end of the school year, measuring how much knowledge and skill students acquired during the year.

NAEP could provide a snapshot of gain scores by testing a subsample of students at the beginning and end of the school year to see how changes vary by place and subgroup.<sup>21</sup> Attainment

of national standards, if implemented, could even be incorporated into the design, through measures of gain and of total knowledge. The former would provide insight into what is happening in the schools, and the latter would provide insight into what the economy, society, community, and family are doing to affect educational progress — both pieces of information that are not yet available but that are critical to raising achievement.

---

<sup>21</sup> This likely would have to be two forms of the same assessment, if it were done with the same students; otherwise, two different samples could be used.



# Facing Choices, Risks, and Difficulties

Considerations exist beyond variability and uneven student performance when advancing toward common education content and curriculum. The nation's inability to implement a program of national standards over the past two decades suggests that efforts are hard to sustain and that reaching consensus is difficult.

## Relationships Between the Call for National Standards and Existing Test-based Accountability

Clear support is emerging for national standards and, for the most part, people want such standards to develop outside the federal government. In the early 1990s, advocacy for curriculum and content standards and for a national standardized test — led by Diane Ravitch, the AFT, the NCTM, and others — called for a more rigorous curriculum and higher-level content. A test-based accountability and sanctions system was not the goal.

Some still support these earlier efforts, favoring standards that are available but optional. Others clearly want a national test, presumably one that accompanies and is derived from content standards. Some want one or both of these to be outside the federal government's jurisdiction and, therefore, voluntary.

Part of the call for nongovernmental national standards stems from disenchantment with the workings of NCLB, the variation in state content standards, and the great differences in the cut-points that states have set for defining proficiency. The warning was explicit in the title of a 2006 Fordham Foundation conference: “When State Standards Go Wrong: Has the Time Come for National Testing?”

The most frequently described scenario would have national standards produced entirely outside the federal government, making them voluntary for states and localities, but encouraged by the federal government. Advocates of such an approach overlap with those who want NCLB to continue, albeit with changes of various kinds — such as establishing common proficiency cut-points among states. But if the idea is to have a set of nonfederal, national standards along with a national test and to incorporate the standards into an NCLB-type of arrangement where the test is required in a sanctions-based system, the test would, by default, become federalized, regardless of how it was developed.

## Assessing the Benefits of Local and State Responsibility

The U.S. education system emerged from the states, and it served the purposes defined by communities through their local school boards or at the state level. The Massachusetts Bay Colony in 1647 established the first effective school law, the purpose of which was to thwart “that old deluder Satan from keeping men from the knowledge of the scriptures.”

Who can deny the legitimacy of localities and states when public schools are still financed heavily by local property taxes and by state taxes? How should people balance local preferences with national concerns?

Over time, the United States has found occasion to take action against inequality of opportunity for poor and minority youth. Two noteworthy examples are *Brown vs. the Board of Education* and the Elementary and Secondary Education Act of 1965. But before people can advocate federal intervention with standards,

they must first consider the history of the origins and uses of education, and whether overriding concerns exist for the well-being of the nation as a whole.

Although there has been a generally continuing consensus around the Elementary and Secondary Education Act of 1965, the federal government afterward lurched from one approach to another, beginning with establishing a cabinet-level agency and then trying to abolish it. And for two decades the nation has made several attempts at establishing standards or goals of one kind or another. What has emerged from this history and experience is that a federal role, if any, must be defined with broad consensus that attains some permanence.

### Recognizing Continuing Conflicts

Algebra is commonly thought to be the same throughout the United States; therefore, national standards for algebra appear, on the surface, to make sense. But a long history of conflict over curriculum reforms in mathematics exists. Referring to the introduction of “new math” backed with federal money after the launch of Sputnik, Michael Kirst, Robin Bird, and Senta Raizen in 1997 wrote: “This time of controversy over curriculum reform sparked by today’s standard setting in mathematics and science is not new. To see how curriculum reform in the United States has long been politically charged and subject to multiple influences, one need only examine the history of school mathematics reform in the 1960s.”<sup>22</sup>

The trio’s paper makes good reading today. A key conclusion is that “winning broad support for content standards is a difficult and inherently political process in which conflicts will arise.”

As recently as March 2008, the question of how to teach math was the subject of the National Mathematics Advisory Panel appointed by President Bush two years earlier. *The New York Times* described the panel’s report this way:

*The report tries to put to rest the long, heated debate over math teaching methods. Parents and teachers have fought passionately in school districts around the country over the relative merits of traditional, or teacher-directed, instruction, in which students are told how to do problems and then drilled on them, versus reform or child centered instruction, emphasizing student exploration and conceptual understanding. It said both methods had a role.*<sup>23</sup>

Math is by no means the only area in which debate has ensued. The wars over reading — especially over phonics vs. a whole-language approach — have been particularly intense and long lasting. Some years ago, a national commission issued a report stating that both approaches were needed and complemented each other. But the debate over the best way to teach reading continued, in one form or another.

Another controversy that has lingered focuses on teaching evolution in the schools. Even in his day, Darwin knew he was stirring up trouble. While aboard the *Beagle*, where he did his work, he said, “I often said before starting that I had no doubt I should frequently repent of the whole undertaking.”

---

<sup>22</sup> Michael W. Kirst, Robin L. Bird, and Senta A. Raizen, “Tensions Between Mathematics and Science Content Standards and Local Politics,” *Occasional Paper No. 5*, National Institute for Science Education, University of Wisconsin, Madison, August 1997.

<sup>23</sup> Tamar Lewin, “Report Urges Changing in Teaching Math,” *The New York Times*, March 13, 2008.

*Evolution: The Remarkable History of a Scientific Theory*, the 2004 work by Edward J. Larson, traces evolution's debate within the scientific community in the 19th century and notes that initial consensus was not reached until the 1930s.

The topic again entered the country's collective consciousness when, in 2005, Kansas' school board removed evolution from state competency tests. And in 2007, a school board vote of four to three settled a debate that had raged in Florida by accepting "compromise language" that identified evolution as "a scientific theory."<sup>24</sup> But the fight could soon begin anew, because Florida law now requires new academic standards in all subjects by the end of 2011.

History is continually subject to revisionism, and even the most current conclusions of revisionists may be unevenly accepted around the country. In the 1990s, Congress debated history standards and, after much contention, voted them down — and that was the end of them.

How desirable is it to elevate the resolution of such strong differences and controversies to the national level, where the nationally organized parties of interest may exert their pressures? Are there ways to set content standards that avoid such basic conflicts? It may be useful to think of our nation as a coil-spring mattress where differences in weight are absorbed and evened out across a larger area. Ultimately, the nation must make a trade-off between accommodating diversity and achieving commonality.

---

<sup>24</sup> Ron Matus, *St. Petersburg Times*, November 6, 2008.

# Creating a National Standards System

A reasonable assumption for now is that national standards will be voluntary. But how can such an ambitious system be created? Recent collaborations among the National Governors Association, the Council of Chief State School Officers, Achieve, the Fordham Foundation, the AFT, the Education Trust, and others have shown some signs of progress.

Since voluntary standards by definition will be unenforceable, sign-on and eventual adoption is dependent on the respectability, integrity, competence, and stature of the individuals involved in the process; and on the perceived soundness of the operating arrangement to plan and create related products.

The source of funding for the initial planning phase also must be impeccable. And the organizations involved must be perceived as open to a careful examination of the problems and possibilities. President of the Thomas B. Fordham Institute Chester Finn lists seven potential pitfalls to achieving national standards, and points to institutional instability as perhaps “the most troubling of all,” saying that there is no suitable place to house standards and take them over for the long haul. He adds that the ad hoc partnerships now assembling to do this job “could fall apart tomorrow” if people die, if organizations change leadership, or if money runs out.<sup>25</sup>

Some have wondered if the federal government might support the operating phase by supplying funds without attaching strings — in other words, by staying uninvolved in the planning so the effort remains national, not federal. For example, federal money assisted the “bailouts” of financial institutions and industries during the economic emergencies of 2008 – 09.

The states and large foundations also could be called upon to sustain an organized standards effort, as they helped fund Achieve and NAEP in the planning stage. With NAEP, in particular, operational funding started at the federal level, entirely through the appropriations process.<sup>26</sup> But a federal legislative base was established and it was quickly recognized how much the states opposed federal encroachment in education.

Because of fear of such federal encroachment, NAEP began by assessing students at ages 9, 13, and 17, rather than at school grades 4, 8, and 12, which came later. NAEP was kept at arm’s length from federal control. The research arm of the U.S. Department of Education opened bids on an RFP for a grant for nonprofit organizations. A NAEP contract arrangement did not come until 1988. For many years NAEP was housed in the Education Commission of the States, which was created by the states. But once the grant was made, the grantee named an Assessment Policy Committee (APC), and this committee elected a chair. In effect, the grantee worked under the APC since the APC assumed full and sole authority over all NAEP policies as soon as it was created.

So NAEP was sheltered from federal influence and flourished while the federal government began to gain more of a role in education. NAEP ultimately became a federal operation under legislation passed in 1988, but its policy remained under control of the National Assessment Governing Board appointed by the Secretary of Education.

---

<sup>25</sup> Chester Finn, Thomas B. Fordham Institute, *The Education Gadfly*, March 6, 2009.

<sup>26</sup> Based on the “general authority” of the NCES authorizing legislation; personal communication from Emerson J. Elliott.

Regardless of funding sources, there must emerge from a widely accepted process an operating organization that, for purposes of this report, this author has called the Standards Entity. Others must accept this group's work as authoritative, unbiased, and uninfluenced by outside pressures. Further, the process by which it is created — and the caliber of the people who guide and staff it — must be respected, for the organization itself will have no true authority.

### What to Do?

This author in no way suggests that he knows, specifically, what such a Standards Entity should do. When the right people converge, they will exchange opinions and find common ground. Further, the agenda must emerge from a respected process if it is to achieve legitimacy.

However, enough can be gleaned from what happened during the education reform era to suggest some strong impressions and conclusions. They are not necessarily in the mold of what have come to be known as content and performance standards of the kind that were incorporated into federal law in 1994 and 2001. Described instead are some approaches toward more commonality in the direction of raising achievement and reducing inequality. Perhaps this discussion will help generate other thinking about the next stage of the standards-based reform movement.

### Learning from Experience

As detailed earlier, the United States has more than two decades of experience with an education-reform plan that began with standards-based reform and morphed into test-based accountability. Federal law has required states to develop and align content standards and assessments, and to establish performance cut-points on a test. The U.S. Department of Education examines these

cut-points, certifying that a prescribed *process* has been followed but not judging the quality of the components. Nonetheless, tests and proficiency cut-points are used for sanctioning schools, to varying degrees, when Adequate Yearly Progress (AYP) targets under NCLB have not been met.

This brings us to the role of the aforementioned Standards Entity, which might first want to review the existing standards system and to study the past two decades of activity. Such an assessment could help suggest a path and would help avoid pitfalls and unwanted side effects, a number of which now have been documented.

Areas of concern and encouragement also have been documented. Several national state-by-state evaluations by AFT and the Fordham Foundation show that existing content standards and tests are of uneven quality. On a positive note, in-depth studies by Achieve typically have been followed by state efforts to correct flaws. But questions remain from careful observers such as Lauren Resnick, W. James Popham, and the Center on Education Policy, all of whom have identified items of concern that could be harmful to teaching and learning. The big question is whether the flaws are correctable and whether they necessarily accompany a sanctions- and test-based accountability system.

Important to the content-standards discussion are the many writings of Popham on how state-level content standards have become wish lists of varying lengths. Committees that establish content standards include subject-matter experts who become advocates of their particular beliefs about what should be in the standards. The group effort to accommodate one another simply grows the list, he says, to the point that far too much exists for teachers to cover in any depth during

the school year. This more-than-can-be-covered content then drives the creation of tests, with too few items in any one content area from which to draw conclusions and with more areas mandated than teachers can manage. The result? Teachers guess at what to teach, hitting it sometimes and missing it other times. This approach is detrimental to instruction and to attempted improvements in instruction. Popham drives home the point with this analogy:

*Please suppose that you are having a dream in which you're standing directly in the middle of a circular-shaped shooting gallery. On the walls surrounding you, there are 100 bulls-eye targets, each about 12 inches in diameter. You've been given a fully loaded gun and told to hit all targets dead center, or as close . . . as you can. However, and here is where the imaginary situation gets bizarre, you are only going to be allowed 20 seconds to do all your shooting.<sup>27</sup>*

Popham carries the analogy further in terms of its effects on instruction. His last point is that it gets worse: “Suppose you’ve just been told that *only half* of the 100 targets will actually count” and that these will only be revealed at the end of the 20 seconds.

Another possibility is that the same test, or a nearly identical test, is used year after year, allowing teachers to become familiar with its content. This approach encourages teachers to narrow instruction to fit the test but has the unwanted effect of score inflation and failure to cover broader subject domain, as well as a resurgence of the “teaching-to-the-test” criticism that has long plagued large-scale testing.

## Finding What's Feasible and Useful

In his new book, Popham asks, “Why is it that the goals we have established for our public schools are so altogether irrational? Distressingly, at the moment that’s exactly what those goals are — completely irrational.”

It is not just that the content standards cover much too much ground. Typically, tests are not aligned with standards, and even when they are, the breadth is too wide and the depth too shallow. So the two wheels of reform working together are defective. In fact, since tests have become the operative elements in the country’s sanctions-based reform system, few people seem to care much about the preliminaries of good content standards and the degree of alignment — both necessary if test results are to have meaning and to measure changes in learning. Two decades ago, content standards were primarily viewed as a means to strengthen the curriculum; now, they are viewed simply as prerequisites for a test.

This primacy of the test, regardless of its validity, can be traced to the first day of the NCLB legislation passed in 2001. NCLB carried forward the 1994 requirements concerning content standards and alignment — preconditions for a test’s validity. At that time, most states were far from fully compliant with these requirements. In fact, norm-referenced tests that compared one person’s scores with another’s were still widely used to determine whether individual students and classes of students reached learning objectives defined by the content standards.

---

<sup>27</sup> W. James Popham, *Unlearned Lessons: Six Stumbling Blocks to Our School's Success*, Harvard University Press, March 2009.



Whatever the status of the standards-based reform system on the ground, the nation passed judgment and issued sanctions using the tests available at the time and by declaring which states had not met AYP in advancing test scores. The quality of the tests did not matter. The U.S. Department of Education continued to tell the states when requirements were not met and urged states to continue progress toward compliance with the law. Even when states met federal process requirements in establishing standards and tests, in-depth examinations found large problems with the resulting substance. As ETS's Henry Braun and Robert Mislevy put it, "A test is a test is a test, and a score is a score is a score."<sup>28</sup>

Since the results of the NCLB tests are not known until the end of the school year, the tests have no role in informing teachers of areas in which students needed help. The tests' role is to judge schools' effectiveness, not the progress of individual students during the course of instruction. In fact, while testing grew by leaps and bounds, the diagnostic tests developed for teachers received little attention, even though a synthesis of studies in the United States and abroad showed that such tests were effective in increasing student achievement.<sup>29</sup>

## Curriculum Standards and "Standardization"

The desire to raise standards and strengthen curriculum content is clear. What is unclear is whether consensus exists for "standardizing" what is taught, as discussed earlier. Most of those who advocate standards do not likely wish to see uniformity of the entire curriculum, of what is

taught in each course, and of pedagogy in teaching approaches. Related questions, then, include: How much teacher creativity should be risked? How much should choice be limited in balancing subjects in a curriculum? More or less history? Literature? The arts? Math?

On this general subject of what to do, and with what impact, this author recommends a 19-year-old article, "National Curriculum American Style: Can It Be Done, and What Might It Look Like?" by Marshall S. Smith, Jennifer O'Day, and David K. Cohen, *American Education*, 1990. These authors detail, in about two pages that are as fresh as if they were written today, four areas to consider in establishing a national curriculum:

- The first concerns how much a curriculum framework should specify. For example, should it specify the particular novels to read? The authors say the NCTM standards represent the flexible end of the framework.
- The second is closely related: the specification of *when*? That is, at what grade level are content and skills required or expected to be taught?
- The third "concerns the depth and breadth of the curriculum; should equal importance and time be given to every decade of U.S. history since 1607, or should more attention be given to certain eras?" Should all areas of science be covered, or just a few key areas?
- Last is "the degree of flexibility allowed to districts, schools, and teachers at the most general level; this might include the proportion of the overall curriculum that is determined nationally," with the remainder to be determined locally.

---

<sup>28</sup> "Intuitive Test Theory," *Phi Delta Kappan*, March 1, 2005.

<sup>29</sup> Paul Black and Dylan Wiliam, "Inside the Black Box: Raising Standards Through Classroom Assessment," *Phi Delta Kappan*, October 1998.



In this country, the authors say, “a national curriculum would have to balance national direction with local discretion.” Although Europe is often mentioned as having a national curriculum that produces good results, it is also identified as having a long history of nationalization. In this country, education started with citizens and localities, percolating up to a serious national involvement only within the past four decades — and entering with force only in the last decade.

### An Alternative to Consider

Is there another approach to specifying rigorous content and matching that content with an aligned test? Considering an alternative is important, for testing in the United States has seldom met that standard. One notable exception is the College Board’s Advanced Placement (AP<sup>®</sup>) program, which has grown in use and popularity. AP curriculum objectives are specified, but its teaching materials are not standardized. Instead, teachers receive a course syllabus and make choices about instructional materials. The outcomes expected are clear, and the test is closely tied to the curriculum. The tests’ utility is evident in colleges’ willingness to accept for credit what each college considers a passing score on each of the tests.

Another College Board program warrants attention here. Pacesetter<sup>®</sup>, as explained by the College Board, is “an integrated program of standards, instruction, professional development, and assessment.”<sup>30</sup>

Achieve led the effort behind a more recent example. Through this work, consensus is growing in a number of states on both the content of an algebra II course and an end-of-course examination. Developments in these areas will be

of great interest to — and have great significance for — the education community.

These approaches rely on a discipline missing in the present content standards: They have curriculum objectives tailored realistically to the time available for teaching a course, as well as a test tailored to match content covered by the course. A Standards Entity could create and supply a national standards test geared to the syllabus for a course, and school systems could create and produce new forms of the same test, likely through contracts with testing agencies.

What remains at issue is the wide variation in student achievement in any one subject at any one grade. This issue is not simple. Colleges have prerequisites for entry into a course. In high school, a prerequisite to algebra II is algebra I. The syllabus for a standard eighth-grade mathematics course, prepared by a Standards Entity, would consider students who had successfully completed a seventh-grade math course. But students would be at quite different levels in their mathematics proficiency, so prerequisites for the eighth-grade course would have to be specified.

One possible way to handle this is to phase in courses that require established proficiency levels, gradually adding them each year until 12th grade. This may sound logical, but it is still flawed because it doesn’t consider the large variation in cognitive development and achievement among students when they *first* enter the public education system.

Despite significant practical and operational difficulties, setting content standards requires more than creating wish lists for schools. Instead, focus should center on establishing realistic goals based on and controlling for the significantly varying degrees of student achievement at any point in time.

---

<sup>30</sup> “Pacesetter” Research and Evaluation Findings, Research Survey, the College Board, RS-06, October 2001.

## The One-Size-Fits-All Question

In the early grades, considerable uniformity of content exists within a subject. Within a school district, for example, teachers are generally clear on what to cover in their second-grade reading classes. By high school, student goals and interests are much more diverse, even as differences in the level and type of student knowledge remain. The long history of efforts to increase the percentage of students attending high school and earning diplomas has culminated in “the comprehensive high school.” A variety of reform movements have tried to change the choices, requirements, and subject-matter offerings over the decades.

The Bush Administration, for example, repeatedly tried to abolish vocational education, which the federal government created in 1917. While Congress did not concur, it did oversee a period of “modernizing” that led to program changes and a new name: Career and Technical Education (CTE).

Underlying these developments is a wave of opinion that all high school students should be prepared to enter higher education *and* to pass college placement tests, allowing them to bypass remedial courses and move directly into college-credit courses. This author contends that high school populations, and the life objectives of those within them, are too varied to have a one-size curriculum. Such a curriculum will not fit all.

Those who advocate a single curriculum have not addressed, to my knowledge, the wide variation in entrance requirements of different kinds of postsecondary institutions, and the

different selection criteria for entrants ready to take credit courses at any particular institution. One size not only does not fit all; it does not even fit all who are going to college.

A Standards Entity promoting a common set of standards in high schools will have to determine the nature of the content it prescribes and the subjects for which it will prescribe content. Decision makers must consider what is best for *all* high school students, keeping in mind that about 30 percent of those who enter high school do not complete it and that three in 10 jobs require a postsecondary credential, per the U.S. Bureau of Labor Statistics.<sup>31</sup>

## Standards for Course Titles?

Having looked at the issues of detailing the curriculum for a specific course, such as American history, and providing an end-of-course test based on that curriculum, this report will now consider *what content a course covers*.

A key recommendation of the 1983 *A Nation At Risk* report was to increase the percentage of high school graduates who take rigorous academic courses. The report specified which courses students should take and how many of them students should take over the four-year period.

Since that report, the percentage of students taking the recommended courses has grown significantly. NAEP, using transcript studies from 1990 to 2005, reported changes in the percent of students “completing curriculum at or above mid-level.”

---

<sup>31</sup> Paul E. Barton, *High School Reform and Work: Facing Labor Market Realities*, Policy Information Report, ETS Policy Information Center, June 2006.

The percentage for White students was 32 in 1990 and rose to 52 percent in 2005. The percentage for Black students was 26 percent in 1990 and also rose to 52 percent in 2005, thus appearing to close the gap. But in reality, the gap in achievement scores by no means closed.<sup>32</sup>

Those in the education community have speculated that more students took courses with the desired titles, but that these courses were merely older ones with name changes, or that the courses were new but the content was watered down, or that the teachers were unqualified. The truth is that no one really knows what happened.

Quite simply, what we know about the content of high school courses does not go far beyond course names. A Standards Entity might devise a course-classification standard that tracks the difference between a “typical” and a “rigorous” course. Perhaps this could be done on a sample basis to establish the dimensions of the variability, followed by a system of spot audits invited by the state or school districts. This could lead, for example, to an algebra I classification ranging from algebra 0.1 to 1.0.

---

<sup>32</sup> C. Shettle, et al., *The Nation's Report Card: America's High School Graduates*, U.S. Department of Education, National Center for Education Statistics, Washington, D.C., 2007. Mid-level curriculum is defined as meeting a standard curriculum (at least four credits in English and three each in social studies, mathematics, and science), plus completion of geometry and algebra II; at least two courses in biology, chemistry, and physics; and at least one credit in a foreign language.

# Getting More Out of Current National Efforts

NAEP has long provided reliable data on the knowledge and skills students possess near the end of fourth, eighth, and 12th grades. But it has not been used to measure what and how much a student learned exclusively during a school year — isolating the effects of a year’s instruction from what and how much a student learned from all sources since birth. Stated another way: *There is a national test of what students know and can do at various points in time.*

Behind the NAEP assessments that span the country is the involvement of many who create frameworks to guide development of the assessment questions and exercises. NAEP’s reach is greater than most realize and much is now gained from NAEP, but even more is *possible* — both in bringing quality assessment approaches to testing systems and in seeing how students do on national as well as local measures. Its scope, in terms of what subjects are assessed, can be greatly expanded.

NAEP’s potential for serving as the desired national test is ultimately linked to what question this country wants answered:

- If the question is how well U.S. students are doing overall and in specific regions of the country, NAEP does that.
- If the question is how well students in each state are doing, NAEP does that.
- If the question is how well students are doing in large metropolitan areas, NAEP does that.
- If the question is how major subgroups of the student population are doing collectively and comparatively, NAEP does that.

- If the question is how much students are growing in what they know as they advance through the school system, NAEP data can be used to do that, as discussed earlier.
- If the question is how well a particular school district is doing, NAEP can do that via sampling — if the district is large enough.

Even today, NAEP works to expand its reach by annually releasing half of its test questions for public use. Testing organizations can create assessments with these items to obtain scores that can provide some comparisons with NAEP scores. (Even though NAEP test items are used, the scores cannot be considered official NAEP scores unless the conditions of the assessment and test administration duplicate those of NAEP.)

Such use of NAEP-released items has a long history with state assessments. In the mid-1980s, when NAEP had only national and regional coverage, the Southern Regional Education Board (SREB) contracted with ETS to construct assessments from released NAEP items. The SREB then administered assessments in three states from 1985 to 1986 and in seven states over the next two years. This was the beginning of a movement to introduce NAEP at the state level.<sup>33</sup> By 1990, the first state assessments were launched under the auspices of NAEP, and state assessment became a regular part of the U.S. education system.

In the late 1980s, the High Schools That Work (HSTW) Consortium, created by SREB, contracted with ETS to construct a NAEP-based assessment with released items. Its goal was to have a national-level standard with which to

---

<sup>33</sup> Personal communication, Mark Musick, January 7, 2009.

compare students in the consortium's participating high schools. To this day, HSTW administers a NAEP-based assessment; now, however, it uses the NAEP framework to develop items.<sup>34</sup>

The state of Florida replicated NAEP reading and mathematics assessments in 1974 and 1975, giving them to a sample of 9-, 13-, and 17-year-olds, and closely matching NAEP's testing cycles, sampling procedures, and other guidelines.<sup>35</sup>

None of these efforts was approved in any formal way by NAEP officials, NCES, or NAGB. But they played a key role in establishing that half of the items developed by NAEP and used in its assessments, as well as NAEP's frameworks, would be available for public use.

On its webpage, NAEP describes the availability of its released items and the frameworks that are developed before each assessment is constructed. It also details how school systems can create content standards through the frameworks, and lists the subjects for which NAEP frameworks are available (<http://nces.ed.gov/frameworkshow.asp>).

Down the road, if educators use NAEP questions in tests that are part of accountability systems, scores might artificially rise because of item familiarity and the practice of "teaching to the test." Claims of comparability with official NAEP scores might be threatened, as a result.

In this era of emphasis on commonality in standards, creating assessments with a feasible degree of comparability to NAEP assessments would be most helpful. If a service to help bring this about were established, it should come from outside the official structure of NAEP.

---

<sup>34</sup> Personal communication, Mark Musick, January 7, 2009.

<sup>35</sup> Personal communication, Thomas Fisher, January 13, 2009.

## Concluding Comments

People's ideas and concepts about national standards differ widely. There is much to be considered when judging the need for standards, and much more must be done so that all of the people participating in the conversation are reading from the same page.

This report offers several ways to view potentially constructive efforts to establish standards at the national level. It provides a brief history of times the nation has started up the hill of establishing national standards and has gone down the other side, largely empty-handed. Additionally, it supplies a brief history and summary information on education standards overall. It does not advocate any particular position but points out pitfalls to be avoided, based on this author's previous analyses of standards-based reform and of the test- and sanctions-based accountability movements.

There still is much work to be done, as ambiguity abounds in the proposals for national standards. For example, whether non-federal — “national” — standards would replace federal involvement in judging schools is just one of the many aspects that remain unclear.

Support for the idea itself also is uncertain, with progress in advocacy of national standards unsteady. Some seem to be backing away from the idea of voluntary national standards at the same time that others, frustrated with seeing state achievement bars set at different heights, view a set of standards that comes from a non-federal source as necessary.

Variation in curriculum rigor is a concern for some who favor national standards. When considering variations in the school system and in non-school experiences, curriculum rigor

presents itself as one of the few factors that could be controlled to minimize wide variations in student performance. But it is unclear how much advocates want to strive for standardizing what schools teach across the United States as opposed to raising rigor in general — which is more in keeping with the nation's history of leaving judgments about education to localities and states.

Another concern is settling issues over pedagogical approaches to teach subjects such as mathematics and reading, and controversial topics like evolution. Specifically, the education community is wrestling with whether to let the differences play out at the local level or to raise them to a national level.

Variation is rife. It's found in the content of courses having the same name, in grade levels in which different degrees of content are introduced, in student scores on achievement tests, and in the height of the achievement and cognitive development platforms from which students are launched into school.

Concerns are rife as well. They include the question of whether national standards would homogenize the curriculum so that all schools teach the same lessons at the same time, as the French do, and whether this is desirable; how far standardization would move school control away from localities; and whether to raise controversial topics to the national level. In a system accustomed to accommodating change, another concern is how to handle new views that may arise — for example, the correct order of subject-matter progression from grade to grade and the evolving objectives in “teaching 21st-century skills,” which has 234,000 Google entries at this writing.

The strong interest in a national-level approach to educational standards will not disappear. The Obama administration and Congress will, of necessity, grapple with reauthorizing the No Child Left Behind Act, thereby ensuring that the issue of national standards remains before the nation.

\* \* \* \* \*

This paper may have clarified matters for some and complicated matters for others. At the end of a symposium conducted by Richard Nathan, he told the story of a person who wrote to him asking for an explanation of revenue sharing. Nathan sent her his book about it. She wrote back: “I’m still confused — but I’m confused on a higher plane.”

The wit of Albert Einstein comes to mind, as he is reported to have said, “You should make things as simple as possible, but no simpler.”

Along these lines, the concept of national standards is complex, but out of necessity.













*Listening. Learning. Leading.®*

*[www.ets.org](http://www.ets.org)*