

## DOCUMENT RESUME

ED 465 821

TM 034 210

AUTHOR Barton, Paul E.  
TITLE Staying on Course in Education Reform. Policy Information Perspective.  
INSTITUTION Educational Testing Service, Princeton, NJ. Policy Information Center.  
PUB DATE 2002-04-00  
NOTE 26p.  
AVAILABLE FROM Policy Information Center, Mail Stop 04-R, Educational Testing Service, Rosedale Road, Princeton, NJ 08541-0001 (\$10.50). Tel: 609-734-5694; e-mail: pic@ets.org; Web site: <http://www.ets.org>. For full text: <http://www.ets.org/research/pic>.  
PUB TYPE Opinion Papers (120)  
EDRS PRICE MF01/PC02 Plus Postage.  
DESCRIPTORS \*Academic Achievement; \*Educational Change; Elementary Secondary Education; Equal Education; Standardized Tests; Standards; Test Construction; \*Test Use  
IDENTIFIERS \*Reform Efforts

## ABSTRACT

This report reviews the experience of the last couple of decades in education reform and the use of standardized testing during that time. It tells of success and achievement and of missteps and problems in the implementation of all the elements of standards-based reform. The report suggests that getting education reform on course involves the recognition that the focus has shifted so much toward testing that some balance needs to be restored, clearly focusing on curriculum and instruction. Many important questions are not being asked about the correct and broader uses of tests. Better quality tests are needed for accountability, and new tests are needed to help teachers improve instruction. An authoritative voice must be created for constructive and responsible testing and assessment data to help organizations dedicated to reform. It is also necessary to create an indicator system that permits tracking progress in each element of standards-based reform. (SLD)

POLICY  
INFORMATION  
PERSPECTIVE

# Staying on Course In Education Reform

by Paul E. Barton

ED 465 821

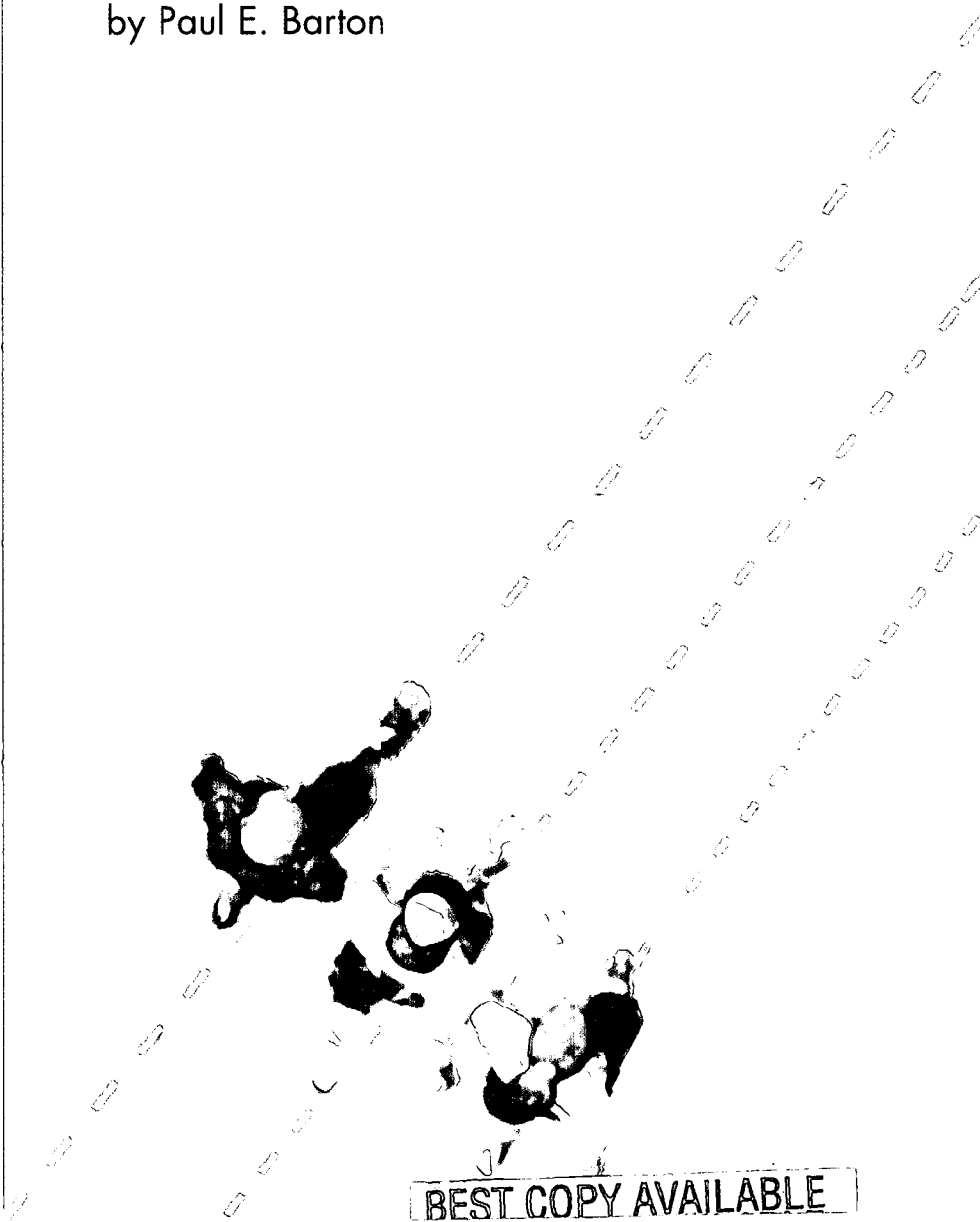
TM034210



*Educational  
Testing Service*

Statistics & Research  
Division

Policy Information



**BEST COPY AVAILABLE**

PERMISSION TO REPRODUCE AND  
DISSEMINATE THIS MATERIAL HAS  
BEEN GRANTED BY

R. Coley

TO THE EDUCATIONAL RESOURCES  
INFORMATION CENTER (ERIC)

U.S. DEPARTMENT OF EDUCATION  
Office of Educational Research and Improvement  
EDUCATIONAL RESOURCES INFORMATION  
CENTER (ERIC)

This document has been reproduced as  
received from the person or organization  
originating it.

Minor changes have been made to  
improve reproduction quality.

Points of view or opinions stated in this  
document do not necessarily represent  
official OERI position or policy.

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

Policy Information Center  
Mail Stop 04-R  
Educational Testing Service  
Rosedale Road  
Princeton, NJ 08541-0001  
(609) 734-5694  
Internet – [pic@ets.org](mailto:pic@ets.org)  
<http://www.ets.org>

Copies can also be downloaded from:  
[www.ets.org/research/pic](http://www.ets.org/research/pic)

Copyright © 2002 by Educational Testing Service. All rights reserved. Educational Testing Service is an Affirmative Action/Equal Opportunity Employer. The modernized ETS logo is a trademark of Educational Testing Service.

April 2002

---

# TABLE OF CONTENTS

Preface .....	2
Acknowledgments .....	3
Introduction .....	4
Asking the Right Questions .....	5
The Forces Shaping Reform .....	7
Two Worlds, Not One .....	10
Finding Thoughtful Ground .....	12
Keeping Our Eyes on the Curriculum and Instruction .....	13
Moving Beyond Quantity to Quality .....	16
An Authoritative Voice for Quality Assessment? .....	19
Summing Up .....	21

## PREFACE

The No Child Left Behind Act, signed by President George W. Bush in January 2002, greatly expands the requirements for setting standards, testing students, and accountability based on these test results. It is an important and ambitious piece of legislation.

Educational Testing Service (ETS) has strongly supported this expansion of the already growing standards-based reform movement. It believes that the use of quality standardized tests can be an important element in efforts to improve achievement. ETS has an equally strong interest in the scope and quality of the implementation of the standards-based reform approach and, more specifically, in the implementation of the new Federal requirements. There are many ways to go wrong during implementation, and there is always a temptation to take shortcuts. The road ahead to a payoff from the new law is a long one, and care and perseverance will be essential.

This report by Paul Barton reviews the experience of the last couple of decades on education reform, and the use of standardized testing during that time. It chronicles a story of success and achievement. But it also chronicles missteps, shortcomings, and incompleteness in implementation of all the elements of standards-based reform. Being aware of this

experience will help inform the implementation of the new Federal requirements.

The report also addresses what is meant by full implementation of the standards-based reform approach. In doing this, it draws from the testimony given by Kurt Landgraf, President of Educational Testing Service, last March before the House Committee on Education and the Workforce, addressing the reforms then being proposed by President Bush. Landgraf set forth criteria for the expansion of standards-based reform; at the present time, in most states, implementation falls considerably short of meeting these criteria.

Obviously, ETS has a vested interest in how these matters play out. It is a testing agency, and it is also chartered to carry out research in education policy and practice. Its Policy Information Center tries to provide useful information to policy makers that can help them in their deliberations.

Drew Gitomer  
Senior Vice President  
Statistics and Research  
Educational Testing Service

---

## ACKNOWLEDGMENTS

As I went through several drafts of this manuscript, I received comments and advice from a number of people who have a wide range of perspectives and experience. I thank them for their help. They include John Barth, National Education Goals Panel; Emerson Elliott, NCATE; Chester Finn Jr., Thomas B. Fordham Foundation; Samuel Halperin, American Youth Policy Forum; John Jennings, Center on Education Policy; Roberts Jones, National Alliance of Business; Richard Kahlenberg, Century Foundation; Glenda Partee, American Youth Policy Forum; Richard Rothstein, Economic Policy Institute; Susan Traiman, Business Roundtable; and Emily Wurtz, National Education Goals Panel.

At Educational Testing Service, this report was reviewed by Ted Chittenden, Dan Eignor, John Fremer, Drew Gitomer, Archie Lapointe, and Harold Wenglinsky.

While many of the above would agree with much that is said in the paper, some would likely disagree with specific points, the emphasis given to some developments, or interpretations of roles played by various sectors. The views expressed are the sole responsibility of the author.

Carla Cooper provided desktop publishing, Martha Mendez and Richard Coley were the editors, Jim Wert was the cover designer, and Trina Black was the production coordinator.

## INTRODUCTION

*In the late 1980s and early 1990s, an energetic education reform movement infused the school curricula with challenging content and built a broad consensus about “content standards.”*

The National Council of Teachers of Mathematics (NCTM) led the way, defining a process to be emulated in other subject areas and in the states. Grants from the U.S. Department of Education sought to extend such work to science, history, and English. Today, a still energetic education reform movement is increasingly centered on standardized testing, where the terms “standards” and “passing test scores” (for schools, students, and, increasingly, teachers) are often used interchangeably. A February 2001 PBS show, *The News Hour*, featured four

discussants on the subject of President Bush’s proposal for testing every student in grades three through eight every year. Two opposed the testing proposal — and seemingly all or almost all standardized testing — for making important decisions, and two were emphatically for it and expressed no reservations about “high stakes” uses of such testing, such as making it the sole criterion of student promotion. It was almost a caricature of the state of the national education debate — Are you for or against testing?

## ASKING THE RIGHT QUESTIONS

*Standardized testing<sup>1</sup> is well entrenched in the public education system, and while the objectives of its use have varied, standardized testing will likely remain — although there are thoughtful people who have concerns about some of its educational impacts.*

The reality is that the issue is not testing or no testing. The important issues have to do with such questions as:

- What *kinds* of tests are to be used to identify failing students, failing schools, and failing teachers?
- Are test scores to be a *sole* criterion, or should there be multiple criteria for decisions that affect students, teachers, and schools?
- How do we set a performance standard when there is such wide dispersion of achievement levels in any one grade?
- How do we create tests, within a reasonable cost, that capture the *full breadth* of the educational experience?
- Are we going to hold a teacher or school accountable for how much math — in

total — a student knows at the end of the eighth grade, or for how much progress a student made *during* the eighth grade?<sup>2</sup>

- How much should we care about the degree to which the test reflects what was actually taught by the teacher? In a standards-based reform approach, how aligned is the test with the content standards? Have the content standards been embedded in the curriculum the teacher is supposed to teach? Is the teacher prepared to teach the new content on which the student is tested?
- How do we make and use tests that actually help the teacher tailor instruction to the needs of an individual student?

There are other questions that could be added, but this list will serve to make the point that the considerations are many and that any debate just on the proposition of whether or

---

<sup>1</sup> Assessment, of course, is much broader than standardized testing. It has been described by Ted Chittenden at Educational Testing Service as including such things as work samples, performance tasks, records of student activities, and teacher-made assessment instruments.

<sup>2</sup> On a statewide basis, only Tennessee measures growth during the school year.



not to have regular standardized testing is too shallow. That debate is largely irrelevant in terms of the choices now on the table — or that should be put on the table. A reasonable reaction of the harried citizen, the generalist educator, or the public official is, “I can’t learn all the ins and outs of standardized testing. I have to rely on the people in charge of the education system to make good tests and use them for the right purposes.” So who are the decision makers and the influencers in the education reform movement?

## THE FORCES SHAPING REFORM

*I*n the abstract, parents and the public are strong supporters of the use of standardized testing.

But parents are often dismayed at how they see tests being used in the schools their children attend. Their support for the broad concept of standards-based reform is intertwined with the practice of standardized testing, the aspect of reform most visible to them.

The political forces — governors, state legislators, and increasingly presidents and members of Congress of both parties — are demanding tests to measure the results of reform or because they believe giving hard tests is itself the reform. The elected officials involved are seldom going to reach a depth of understanding that permits them to make sound judgments about the appropriate uses of tests. To the extent that political officials determine who runs the schools in the states, they will find people to appoint who are on the fast-moving testing train.

Professionals in the education system — superintendents, principals, and teachers — have a great many concerns about how standardized tests are being used in standards-based reform, often thinking that the tests are being misused, based on their professional knowledge of education. A sizable proportion supports testing as a useful part of instruction and using tests for purposes of accountability, but many such supporters also see flaws in the way testing is being done in their state, district, or school. Mostly, the people who stoke the fires that drive the fast-moving

testing train dismiss the professionals as trying to avoid being held accountable. Of course, what many educators have not faced up to is that the information now available about student performance is not considered credible. Providing other credible information about how schools are faring might relieve pressure for more and more testing.

The business community has been an effective force in the education reform movement of the past couple of decades and has recently begun to advocate improving the quality of teaching and teacher pay. For example, the recent statement by the Business Coalition for Excellence in Education (February 26, 2001) is a blueprint for a responsible approach to standards-based reform.

But the system that is actually developing in a large number of places does not follow that blueprint, particularly in the matter of the appropriate role for standardized testing. The business community has pushed the use of tests for accountability, but it has not looked closely at how testing is being done in practice or whether the basic principles of testing are being followed or violated, nor has it followed through to urge specific corrective action. If any force can be effective in getting standards-based reform back on course, it will be the business community. It has a compelling interest in raising student achievement, and it has the ability to command attention.

Then there is the testing community. It includes the companies that make tests, the officials in the education system who establish and run the testing systems, the educational measurement experts and psychometricians in universities, think tanks, testing organizations, and the staff or members of such respected organizations as the National Research Council, the American Psychological Association, the National Academy of Education, and the National Council on Measurement in Education. These are the people who establish and guide the profession of educational measurement. In their standards, in their principles, publications, and admonishments, they make available a body of knowledge that could permit the establishment of a state/district/school testing system that would be both responsible and useful. They don't all sing in the same key, but they mostly sing from the same hymnal. However, their impact on actual testing practice is limited.

This is not to say that the experts on testing should be in charge of policy any more than doctors should be in charge of health care policy or soldiers in charge of military policy. Testing experts have information that should be heard by people who do make policy, people who make judgments about what to do based on that information and other, broader considerations. But such experts in testing do not necessarily know how best to use tests to improve instruction and learning.

The companies that make tests may well be aware of the limitations of their tests, and their literature will likely be clear as to what their tests measure and what standards they have applied to establish validity and reliability. But many tests result from a decision by a state to put out a Request for Proposal (RFP), on which testing companies bid, and the RFP specifies in detail what the state wants.<sup>3</sup> The state owns the resulting test and uses it in any way it pleases.

Testing companies also make and sell or administer their own tests, but to do so they must sell what the market demands. Although there are many exceptions, too often that demand is for cheap-to-administer standardized tests that can be used to rank students and schools for purposes of accountability. There is a lot of rhetoric about using tests to “diagnose” student needs and inform teachers about how to help individual students. But there is no large market here, and test development and validation would be expensive. The rhetoric we frequently hear about using test results to improve instruction does not match the actual uses being made of tests. The tests we now use don't tell teachers how to correct their approach to an individual student; they are typically given at the *end* of the year and are used to pass or fail students or schools.

---

<sup>3</sup> A state also may develop its own test and tie it closely to its educational objectives.

So the power and responsibility for standardized testing is diffuse. No group or sector steps forth — in any systematic way — to combine testing knowledge and practice to keep actual test use within the confines of the tests' capabilities or to bring together test use with sound educational practice. Rare are events such as the one in the early 1980s when Gregory Anrig, President of Educational Testing Service, told then Governor Bill Clinton that Arkansas could not use the National Teacher Examination® on people already teaching because that was not a valid use of this test.

## TWO WORLDS, NOT ONE

*There is another huge complication in the important enterprise of harnessing standardized testing in elementary and secondary education for the service of teaching and learning.*

Testing, with its theories of “mental measurement,” grew up largely as a separate enterprise from instruction. This separateness was reinforced by virtue of testing’s early preoccupation with measuring “intelligence,” by its use for sorting students into tracks or ability groupings, and by its gatekeeping role in admission to colleges and graduate schools. The role of standardized testing in standards-based education reform and in measuring achievement relative to established standards is relatively new. This kind of testing does not come from the educational establishment, its scholars, its pedagogy, or its schools of education. It comes from this community of mental measurement and is joined with instruction in the classrooms in a shotgun marriage.

Not surprisingly, this separate evolution of mental measurement and educational pedagogy often results in different images of how the two relate in advancing learning. A “learning theory” that encompasses both has not really developed. There are even different views of what constitutes a standardized test that claims to be a valid measure of achievement in the classroom, which were recently described by Robert E. Stake of the University of Illinois, who has spent many years in both evaluation and psychometrics. (See his address at the American Psychological Association meeting in San Francisco, August 24,

2001, titled “Evaluation of Testing and Criterial Thinking in Education.”)

So, side by side, competing views often exist. There is the view that a standardized test should tell us whether the students are learning what is being taught, alongside the view that the imposition of a “hard” test is the way to change the curriculum.

There is the view that such a test is simply an estimate of what the student knows at a point in time, to be used with other indicators of student progress in making important decisions, alongside the view that one test can show what the student knows well enough and, in some places, how effectively the student was taught.

There is the view that a standardized test can encompass the full range of effects of what the teacher is supposed to accomplish, right alongside the view that a few books in the library (a few test questions) can hardly represent the whole of what a library has to offer, and that the efficiencies required in test making and administration are bound to restrict how much the test captures of the total learning experience.

There is the view that a single test can be used for multiple purposes, alongside the view that a test to evaluate schools is not also a test for diagnosing individual student needs.

Even within the testing and education community we find different conceptions of the purposes and uses of testing in the service of raising student achievement, so it is no wonder there is a lack of common understanding. And the other actors in the education reform play are even further removed from a common educational philosophy and theory of educational improvement that encompasses standardized testing.

## FINDING THOUGHTFUL GROUND

*Out of this balance of power, of pressures to improve achievement, of misunderstandings of what testing can and cannot do, of impatience with the equations and formulas of the measurement experts, of the demand for quantifiable outcomes from the business and political communities, of the decline in power of the professional educators as they are declared responsible for the U.S. being “a nation at risk,” has come an education system ruled increasingly by standardized tests.*

Such testing has taken on an aura of inevitability, and as the testing train gathers speed, few are willing to step in front of it.

The problem is not so much that standardized testing has limitations in the role of sole arbiter of what is happening to the quality of education. The problem is that so little ground remains for being thoughtful about the use of testing, about more complex alternatives for making critical decisions about students, teachers, and schools. It isn't just that misuse of tests leads to bad judgments and decisions; it is that we are denied the benefit of using tests in ways that can be supportive of teaching and learning. There is a real

danger, because of an eventual backlash against the excesses, that we will retreat from developing testing into what it has the promise of becoming.

What we need to do is explore the ways in which testing can support instruction and make the investment necessary to bring such testing online — literally. As we move toward the greater use of computers in delivering instruction, we should build assessment in. We need testing that can give regular feedback and guidance to the student and the teacher, whether through traditional instruction or through computers.

## KEEPING OUR EYES ON THE CURRICULUM AND INSTRUCTION

*But the problem farther around the corner is not just that we are bypassing the opportunities to develop testing that can help in instruction, or that we are using tests in ways they have not been validated to be used.*

The problem is that the serious discussion of educational reform has turned primarily to *how much we are testing and whether we are being tough enough in holding educators' feet to the fire*. That is primarily what the media are reporting. That is what candidates for office use as their lead. In too many places, testing is becoming the treatment rather than a means of finding out whether the treatment is working or of providing teachers with information designed to tell them how to improve instruction for an individual student. Good testing has its constructive role to play, but what our attention should be riveted on is the curriculum,<sup>4</sup> whether teachers are prepared to teach what we want the curriculum to be, and whether teachers are actually delivering that curriculum in the classroom.

In the late 1980s and the early 1990s, as noted at the outset of this article, education reform began on a track of getting agreement on what students should know and be able to do. It first took the form of developing "content standards" in mathematics by professional educators in the National Council of Teachers of Mathematics. These were to be the guides for changing the mathematics curriculum in the classroom. The need to improve the math curriculum was confirmed in the last

international mathematics and science assessment (TIMSS), which showed that American students did respectably in the fourth grade, but lost ground relative to students in other countries between the fourth and eighth grades. Compared to a detailed analysis performed of the curriculum used in those countries, the curriculum in American schools showed a lack of focus, lots of repetition, and a lack of depth. The conclusion was that we must change the curriculum. To raise achievement, we must:

- agree, at the appropriate government level, on the "content standards," what students should know and be able to do;
- translate these standards into new curriculum, and secure the instructional materials necessary to deliver it in the classroom;
- prepare teachers to use the materials and deliver the new curriculum;
- monitor the delivery of the curriculum by the teachers, through a variety of means, such as checking lesson plans, administering questionnaires, and observing classrooms.

---

<sup>4</sup> The term curriculum is used as shorthand for a focus on instructional content and the methods of delivering it.



The role of testing in raising student achievement is a multifaceted one. We should start with helping teachers develop their own tests to get feedback as part of their instruction (little of this is taught in schools of education). We need to bring new tests to the classroom — at the *beginning* of the school year — that provide information to the teachers about student weaknesses in knowledge, as a guide to teachers to help students overcome those weaknesses. Standardized tests, aligned with the delivered curriculum, can then be used to see if students are learning what the teacher is trying to teach them. The results could be a factor in assessing the student and in holding teachers and schools accountable for results.

This is by no means intended to relegate testing to a minor role. What many policy-makers see in testing is something that will gain traction for change, something that will have to be confronted. Good data on student performance can focus attention and, where there are consequences, require that something be done to raise achievement. What we want to avoid is unintended consequences. When the education system begins to perform better and achieves better credibility in the policy community, the drive for testing will lose some of its urgency.

If “no child is to be left behind,” what we need for every year and for every grade is to know whether the curriculum being delivered in the classroom reflects the agreed upon “content standards.” And we need to know the status of teacher preparation to competently

deliver that curriculum. Let us have the detailed reports, and let us be informed by a debate over how far we have gone and how far we have to go in getting a rigorous curriculum into the classroom.

Part of knowing whether the desired curriculum is being delivered is knowing whether the teachers are being prepared to teach it. While there are good models of teacher preparation, the effort falls far short of the need. In March 2001, the results of a survey conducted by the National Staff Development Council were released. The Council’s members are curriculum development coordinators, school administrators, and other educators. Of the members polled, barely half indicated they were in districts where state standards were used as a basis for designing staff training, and many were in school systems that made a limited investment in any kind of professional development. Knowing whether teachers are prepared for the new instruction required by new standards is critical information for tracking the installation of standards-based reform.

Developing and reporting good indicators of the implementation and success of standards-based reform is not a simple matter. But who would have expected it to be simple to raise significantly the educational achievement of the American student? The standardized test is a simple approach to a record-keeping system; it provides an easy-to-understand count of how many pass and how many fail the test. The standardized test has become the favorite shortcut to raising student achievement, but it could become a shortcut across quicksand. Testing has

considerable promise as an important tool in quality education. It would be very unfortunate if its promise sank from the weight of the burden placed on it in the mistaken belief that we can bypass the difficult work of changing the curriculum, developing and securing better instructional materials for the classroom, preparing the teachers, and developing quality indicators for tracking progress.

All this requires that we be thoughtful and tolerant of the complexity of the actions required to raise student achievement. We need people in positions of power willing to listen and learn about what testing can and can't do in the reform process. Dealing with complexity, ambiguity, and uncertainty is what we expect in a highly educated citizenry, and these qualities will be required if we are to achieve standards-based reform as a means of raising the levels of knowledge and improving the critical thinking of a new generation of citizens.

## MOVING BEYOND QUANTITY TO QUALITY

*But what is the place of standardized assessments in standards-based reform?*

While this paper has described a number of constructive roles for testing, such as enabling teachers to make better tests and to use diagnostic testing at the beginning of the year to help teachers improve instruction, there is a clear role for end-of-year testing for accountability purposes. This is where the testing movement has been going, and it is no longer in its infancy. It has now “gone to scale,” is beyond adolescence, and has entered the teen years of vigor and confidence. There is no frailty here; the large-scale uses of testing for accountability will not be altered now by concentrating on quality as well as quantity. A couple of years ago, when then Secretary of Education Richard Riley called for a “midcourse correction,” the response was to circle the wagons, cite polls to show public support for testing and accountability, and point to all the good things about testing. Public and political support for testing remains strong; it will not be endangered by improvement.

If there is a danger to the widespread use of testing for accountability, it will most likely be in ignoring the problems of quality and appropriate use of testing in a total system of standards-based reform. As right as it may be to use tests in the reform effort, there are many ways to go astray in constructing and implementing testing systems — and in much

more basic ways than what typically reaches the newspapers, such as instances of machine mis-scoring or test questions with wrong answers. While there is room for disagreement as to what constitutes best practice, there are testing practices in widespread use that are fundamentally flawed. Unfortunately, the national record keeping, state by state, as to what is being done in the accountability systems is very primitive, and the need for better indicators has been described above. But there are a few things we do know and should be alarmed about.

- In the 1994 amendments to the Federal Title I program, content standards, performance standards, and aligned assessments were mandated. But the Department of Education has found that only about half the states have performance standards that would permit the test to be aligned with the state content standards. Such alignment is a starting point; otherwise, the students are not being tested for what the state is requiring them to know. Only 17 states had been granted full approval by November 2001.
- In about a third of the states, norm-referenced tests are being used for accountability. These tests rank students by percentiles and can be used to rank schools. But they do not measure how many students are

reaching a specific standard, such as being “proficient” in mathematics. These tests tend to be national commercial tests and may not match the content standards of a particular state. Even where they have been customized, they do not measure well who reaches a particular level of achievement because of the way questions are selected to permit ranking; questions that all or most of the students can answer would be thrown out, for they do not help distinguish one student from another.<sup>5</sup> All states need a test that is designed to show how many students are reaching a specified level of absolute achievement.

- In an in-depth study being carried out in four states under the auspices of the Council of Chief State School Officers, the findings were as follows: “Alignment between assessments and standards varied across grade levels, content areas, and states without any discernible pattern.” (The study was conducted by Norman L. Webb and is titled *Alignment of Science and Mathematics Standards in Four States*.)
- A study is in progress in 11 states under the direction of Anthony Bryk, of the University of Chicago, titled *The Survey of the Enacted Curriculum Project*. In mathematics in Grade 4 and science in Grade 8, the findings were that “less than half the intersections of content topics [as reported

by the teachers]...were in common with the assessment items found on the state...test.” In other words, half the time the students are being tested on things they were not taught.

- In many instances the test score alone is used in important decisions about schools and students. No testing company or testing expert this author is aware of agrees that this is sound practice or proper use of standardized tests of the kind that are generally in use today.
- In what may be an extreme example, several years ago a state moved from its state content standards almost directly to a high stakes test based on those standards. The result: 97 percent of the schools failed the test. The curriculum had not been changed to reflect the state’s standards that were used as specifications for the test. At the other end of the spectrum, there are states that have made the effort and taken the time to proceed in an orderly sequence in the implementation of standards-based reform.
- In November 2001, the American Federation of Teachers issued its latest round of state-by-state evaluations of standards-based reform implementation. It concluded that
  - almost a third of the states’ tests are based on weak standards;

---

<sup>5</sup> Norm-referenced tests may vary in how well they measure achievement against a standard. However, “criterion” referenced tests are clearly better designed to do this.

- forty-four percent of those tests are not aligned to the standards;

- fewer than one-third of the tests are supported by adequate curriculum; and

- one-third of the tests used in decisions regarding promotion or graduation are not aligned to the standards (these conclusions are quoted from *Making Standards Matter 2001*).

By what criteria, then, do we establish a standardized testing system as a component of an accountability system? In his testimony last March before the House Committee on Education and the Workforce, Kurt M. Landgraf, President of Educational Testing Service, offered this advice on "How to Do It Right." Among his criteria are the following:

- continued development of unambiguous standards in each state that the education community and the public accept as meaningful;
- state curricula that are linked to state standards;
- instructional materials that are linked to the curricula;
- professional development for teachers and administrators to understand the standards,

know the curriculum, and skillfully use the learning materials;

□ the opportunity for all students to learn the curriculum's material;

□ prior notice to students of testing requirements;

□ assessments [that are] linked to standards....

Careful review of the 50 states would reveal that all of these conditions seldom prevail. In fact, a case has been made by W. James Popham that most of the tests now in use for accountability are not really measuring what students learn in school, that the present standardized achievement tests are "misnamed and misleading." Popham has long been a leading authority on testing in the United States, an author of more than 20 books, and a proponent of the development and use of standardized testing. As to the question of whether it is possible to build standardized tests that accurately measure what is being taught in school, he replies, "Absolutely. But those tests need to be built with a specific role in mind. We need to evaluate a school based on how much students have learned in that school. But we'll never do so if, because of misunderstandings about the role of traditional standardized achievement tests, we continue to use the wrong tests when judging our schools."<sup>6</sup>

---

<sup>6</sup> W. James Popham, emeritus professor, University of California, writing in *Education Week*, 21(3), p. 46, September 19, 2001.

## AN AUTHORITATIVE VOICE FOR QUALITY ASSESSMENT?

*W*e've talked about what should constitute standards-based reform, how testing fits as an accountability element, how curriculum and instructional content has slipped from its central focus at the beginning of the reform movement of the late 1980s, and how we need better tracking of all the elements of reform.

Let us now return to where the discussion started, with the lack of development of broader uses of standardized testing and the need to develop assessment to further student achievement beyond narrow uses for accountability.

There is no clear and easy way to bring about change in the testing enterprise. First, we must achieve a broader understanding of how to interrelate instruction and responsible standardized testing and how to use testing in the service of teaching and learning. Much of what is going awry is the result of lack of understanding and knowledge on the part of those in positions of power and influence. This perception is based on the observation of how often testing is falling short of its potential after a dozen years into the reform movement. Of course, many of the questions raised in this paper are already on the radar screen of important organizations and individuals working to improve student achievement; it is not a barren landscape.

Unfortunately, what those with the power and influence to promote testing often hear are objections to uses of tests from those with an anti-testing perspective and from those who oppose tests having any role in education improvement. What is needed is a voice that advocates the constructive and responsible use

of tests, a voice that points to positive uses — and clear examples from experience — as well as instances where testing is now missing the mark and where tests are used wrongly in accountability. It needs to be a voice that will be respected among those leading the education reform movement. It needs to offer more than technical and scientific information about testing; there are quality sources of such information now, such as the university-based Center for Research on Evaluation, Standards, and Student Testing (CRESST), and the National Research Council. It needs to be from a source that is free of the charge of promoting a self-interest. And it needs to be from a source that is not seen as promoting a particular ideological viewpoint in education or espousing one side or another in the polarized debates that are taking place over testing issues.

There are organizations and sectors that, given such an authoritative source of knowledge and information, could put it to good use. It could be put into play by business organizations such as the Business Roundtable, the National Alliance of Business, and the National Manufacturers Association or by education associations, teacher unions, education committees in state legislatures, and governors or by national leadership organizations such as the Education Trust, the National Governors

Association, Achieve, the Center on Education Policy, and the Institute for Educational Leadership, to name a few. While there are often informed and enlightened views in such organizations, no one of them is likely staffed well enough to provide in-depth coverage of these matters. (Of course, almost any one of them could conceivably create a capability to fill this role.) But, some source with adequate resources and credibility is needed that can serve all of them.

The way such things get done is to have the need for an “authoritative source” to be recognized, say, by the now almost institutionalized Education Summit gatherings that have recurred since the 1989 Summit in Charlottesville, Virginia. Then, that “authoritative source” needs to have financing that does not associate it with one viewpoint, possibly with foundation involvement and contributions from different sectors of the education reform movement. And, it needs to have a leader who has achieved the kind of recognition that engenders trust and respect — not necessarily someone who is a

testing expert, but someone who would bring together the needed expertise.

The channels to good practice through the national organizations should not be over-emphasized, however. Each state needs a source closer to home, such as a policy center in a state university, but all could draw on a central resource.

For education, these are times of great expectations, broad political support for reform and improvement, continued attention of presidents and governors, and strong faith in large scale use of standardized testing as a component of what has become “standards-based reform.” The promise is real, and the prospects, hopeful. Now we need to draw on the experience of the last couple of decades, distill the accumulated knowledge of those whose careers are devoted to teaching and research, make sure we are putting into place *all* the elements required for success, and have the patience to do testing right. Staying on course is likely the only sure way to a higher achieving society.

## SUMMING UP

### Getting *on course* involves:

- Understanding that the focus of education reform has shifted so much toward testing that some balance needs to be restored, clearly focused on improving curriculum and instruction;
- Realizing that a lot of important questions about the correct and broader use of tests are not being asked;
- Making better quality tests for accountability, using them correctly, and, at the same time, developing new tests for the purpose of helping teachers improve instruction, recognizing that there is no such thing as an all-purpose test;
- Recognizing that the base of support for testing and accountability has become strong enough that it can endure a new effort to weed out bad practices and improve quality and that, indeed, its future support may well depend on it;
- Creating an authoritative voice for constructive and responsible testing and assessment that can help organizations dedicated to education reform;
- Creating an indicator system that permits tracking progress in *each* element of standards-based reform.

Many organizations and power centers these days are, one way or another, putting education institutions to the test. But they, too, are being tested, and when the results are in and the tests are graded by historians, we will know whether they are to be recorded as facing the challenge in the early 2000s of identifying constructive and responsible uses for standardized assessment and achieving effective reforms. There are organizations that are mindful of the full range of efforts needed for effective standards-based reform and that are working hard and responsibly to gain implementation. But the separate development of the testing movement, the highly technical nature of the standardized testing enterprise, the relatively recent introduction of such testing in elementary and secondary education for purposes other than ranking and sorting students, and the lack of readily available knowledge of models of good practice actually in use have handicapped the overall reform movement.



There are missing elements in these reform efforts such as attention to quality in testing, focus on a rigorous curriculum, and an authoritative and knowledgeable voice on effective testing practices to which the organizations involved can turn for help. Also missing is a record-keeping system on *all* the elements that constitute a full standards-based reform effort, so we know where each state and district stands and whether steps are being missed and shortcuts taken. Filling in those missing elements would put us back on course in education reform.



**U.S. Department of Education**  
Office of Educational Research and Improvement (OERI)  
National Library of Education (NLE)  
Educational Resources Information Center (ERIC)

TM034210



## NOTICE

### Reproduction Basis



This document is covered by a signed "Reproduction Release (Blanket)" form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.



This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").

EFF-089 (3/2000)