

The Elementary and Secondary Education Act and Standards-Based Reform

By Paul E. Barton

Many considerations are involved in the reauthorization of the Elementary and Secondary Education Act of 1965, including issues on which there is debate and disagreement. This paper addresses only one aspect of the 1965 Act: the 1994 amendments regarding the requirements that states adopt “standards-based reform.” Specifically, it discusses the requirements that states have content and performance standards for all students and that tests be aligned to those standards.

I. VALUES AND PRINCIPLES.

Standards-based reform was a major effort for raising educational achievement in the 1990s, but it began earlier. It traces back to the standards rolled out in the 1980s by the National Council of Teachers of Mathematics (NCTM), to the Governors’ Education Summit in 1989 and the resulting National Education Goals Panel, to concepts of “systematic” reform, to the Goals 2000 legislation, and perhaps most importantly to the 1994 reauthorization of the Elementary and Secondary Education Act, particularly its Title I component. That title required the states to have a triad of content standards, performance standards, and aligned assessments for all students and for this model to be applied to Title I recipients.

The general approach of standards-based reform has garnered a wide consensus. In terms of values and principles, the triad suggests an undeniably logical progression: Get broad acceptance of what students should know and be able to do,

decide how well and how much of it they must master, align standardized tests to the content, and use the results to see if students are reaching the standards.

Although there are many differences in values and opinions about how to raise student achievement, there has been little quarrel with the underlying logic of this formulation of how to approach reform. Perhaps that is because it describes a process, not the means by which achievement is to be raised. It seems to this author that the basic direction set in this approach can continue to serve the goal of raising student achievement. But there is great need to take stock of how this approach has played out in practice, with a view toward steering it back to the broader visions of its creators.

Standards-based reform in practice has come to be something often quite different from the approach espoused by mathematics teachers that caught the nation’s attention in the mid-1980s. (In

fact, in the early stages of reform, Assistant Secretary of Education Diane Ravitch tried to spread this kind of effort by education practitioners to other subject areas, by providing grants to comparable groups to do this work.)

What I argue in this paper is that a funny thing happened to standards on the way to the classroom. The movement needs to be put back on the track it started from, and Title I is a critically important place to do this. That's because Title I is driving a lot of what is being done in states that did not take this initiative on their own or that altered the approach they started with in order to comply with the new legislation.

II. THE LESSONS WE ARE LEARNING.

The period of experience for standards-based reform begins from about the mid-1980s, as NCTM standards were coming to fruition. I will summarize what I think we are learning from this experience and why what we are learning should give us pause. Also, I will point out what we don't know and haven't been keeping track of, and how that may be hurting us a whole lot. We are, I believe, at a point where we need to establish higher standards for judging when standards based-reform efforts are doing what they were originally designed to do.

A. A Standard Is Not a Test and a Test Is Not a Standard.

In the initial conception of standards-based reform, the idea of high standards was to define challenging subject matter that teachers were expected to teach and students were expected to master. It has come to pass, however, that the term "standards" is used almost interchangeably with standardized tests. How this came about is somewhat of a puzzle, but one runs into it all the

time. A typical expression is Yes, we are establishing high standards; we have a new hard test and we have accountability.

Webster's New Collegiate Dictionary gives this definition of standards (among several, of course): "the fineness and legally fixed weight of the metal used in coins." A scale is what is used to measure the weight, to see if the standard has been met. But the scale is not the standard. And a test is used to measure whether the educational standard is being met; it is not itself, the standard.

Having higher standards means doing what NCTM did. In fact, NCTM did not even develop a test. The push to get standards in place is moving too quickly to a test to hold all parties accountable. Testing is what should be done at the end to find out if standards-based reform is working. The test needs to be placed in proper perspective, as one piece of the total effort.

B. The Establishment of "Content" Standards Is Just a First Step in a Necessary Sequence of Events.

The establishment of content standards has resulted in having in the state capitol a document, likely a thick one, describing what students are expected to know and be able to do in all subjects and in all grades. This is not a "curriculum." And it is even further removed from being a set of lesson plans. It cannot even be translated into a curriculum until a decision is made as to a balance among all the subjects for which content standards have been created. Content specialists in American history, for example, will want the coverage and depth that they think appropriate for the subject. But how is the time to be allocated among all the subjects?

And, for each subject, how are these content standards to be translated into a curriculum for American history in the 10th grade? Who is responsible for doing it: the state chief state

school officer? The district superintendent? What is the authority and responsibility of the local school board?

One has to assume that the curriculum must change to meet the new (higher) content standards. That may mean changing the textbooks, the workbooks, the software for computers, and other instructional materials. And all this change has to be translated into the day-by-day lesson plans used by the teachers. Moreover, the teachers have to be instructed in what the new curriculum is, and in how to deliver it. But the way things have developed, content standards are being treated, in some places at least, as ends in themselves, merely as a step in getting to a standardized test. If Title I is to continue to drive standards-based reform, it needs to come to grips with what has to happen after the content standards are developed.

C. The Role and Construction of Performance Standards Remains Ill Defined and Not Well Understood.

In the original formulation of standards-based reform, the term “performance standards” seemed straightforward. The content standards were what students should know, and performance standards were how much of this they needed to show that they had mastered. But how were performance standards to be derived? The major accomplishment at the time was the rolling out of NCTM standards; but these addressed content, and not a way of determining a standard in terms of the amount of the content the students would be expected to master, or the depth of their knowledge and competency. The truth of the matter is that this was uncharted territory, and not the province of existing professions.

Content standards were the creation of educators, in the case of NCTM standards. And while the participants were often broadened to

include others in arriving at standards documents at national and state levels, they were all concerned with the substance of what should be taught in classrooms.

The professions of educational measurement and psychometrics deal with testing and assessment. They know how to move from content standards to specifications for a test. They also draw upon developed practice in setting passing scores, or “cut points,” that students have to reach on a test. Some accepted techniques exist for doing this in which panels of experts look at the individual test items and decide how many a test-taker has to get right in order to meet some agreed-upon level of competence (the frequently used “modified” Angoff method is one of these). But the intervening step of establishing “performance” standards usually gets by-passed, and no professional body has stepped forward to tell us how best to link content and performance standards.

The matter of the gap between content and performance standards was examined by a leading educational measurement expert, Bert Green, at Johns Hopkins University in a lecture he gave at Educational Testing Service. He summed up the situation this way (from *Setting Performance Standards; Content Goals and Individual Differences*, ETS Policy Information Center, 1996):

“The performance standards have to reflect the content standards. The bridge from the content standards to the performance standards depends on the test specifications, the item writers and test editors, and on the resulting performance measurement scale. Logically, it would seem preferable for the judges to set standards just on the content domain. They could identify what parts of the domain are basic, what parts go with proficient persons, and what parts would mainly be mastered by advanced students. It is

not at all clear how to do this [emphasis supplied], but a way might be found. Judges might also be useful in evaluating the bridge from content to performance. This would seem a more straightforward task than imagining the behavior of marginally competent test takers...”

To the extent that we continue to simply look at individual test items to set passing scores on a test, we do not have a known and recognized process to develop performance standards that mediate between content standards and what students have learned. This is a piece of unfinished business in the development and application of a standards-based reform approach.

D. A Way Needs to be Established to Monitor the Extent to Which the Content Standards Are Being Delivered to the Classroom Through the Curriculum.

None of the indicator systems we now have in place provides a vital piece of information. The content indicators may look good. But is the state changing the curriculum and preparing the teachers for it? And if the states say they are, how do we know what is reaching the classroom?

Some would answer that this is what we find out when we give a test that has been aligned to the content standards. But student performance on such a test is not a clear indication of how closely the instruction is in line with the content standards. If students do poorly it may be because the content did not reach the classroom. It also may be that the instruction was aligned but the instruction was not competent. Or it may be that the curriculum was delivered but the learning environment in the school and/or at home was not conducive to achievement (for example, there may be problems with discipline and orderly classrooms). It may tell us that while students

progressed a lot from where they were when they started, they started from such a low level they could not reach a common standard in the same amount of time as students in a different school who started from a higher level. (This might be because they had poorer instruction in earlier grades, or because they came from poorer neighborhoods, or because of fewer opportunities for learning experiences in early childhood).

This matter of translating state content standards into the delivered curriculum becomes very important if we are to know how much standards-based reform is actually changing instructional content and delivery. It is also very important in how we use tests for student, teacher, and school accountability. Is the teacher responsible if the curriculum he or she is provided is not changed? Is the student responsible for meeting standards not taught? And how can we tell from the results of the test alone whether the curriculum delivered in the classroom had been changed to reflect content standards? We are putting too much of a burden on the test, and we will not get the information we need from it.

But the overriding concern is that we may or may not be getting the job done when we check off the box that says “content standards,” which is what the state must do to comply with the requirements of Title I. What we want is to have a rigorous curriculum in the classroom and have it delivered by a teacher who is prepared to deliver it.

E. Alignment.

Alignment has been referred to several times in the above points. The term is currently used to refer to an alignment of the test with the content standards, and sometimes with the performance standards (which I have said do not have an agreed-upon meaning that we know how to make operational). Determining such alignment is not

a simple proposition. A lot of work has been done recently in developing a rigorous approach (see the Council of Chief State School Officers' publication *Alignment of Science and Mathematics Standards in Four States* by Norman L. Webb). A summary of findings is as follows (from the report's Executive Summary):

“Reviewers analyzed the alignment of assessments and standards in mathematics and science from four states at a four-day institute.... Data from these analyses were processed and used to judge the degree of alignment on four criteria: categorical concurrence, depth-of-knowledge, range-of-knowledge correspondence, and balance of representation....

Alignment between assessments and standards varied across grade levels, content areas, and state without any discernable pattern....

A goal of this study was to develop a valid and reliable process for analyzing the alignment among standards and assessments.”

Having such a requirement in federal law puts a lot of responsibility on administering staff to make these analyses and render judgments. But the considerations go beyond the matter of alignment between the content standards and the test. What we really need to know is the alignment between the test and the curriculum. Whether the curriculum does or does not reflect the content standards, we want to know about the alignment of the test to what is actually taught. It seems reasonable that we should hold students accountable only for what it is they have been taught.

We may also want to hold teachers accountable for teaching what they are supposed to teach, and ways need to be developed to do this (evalua-

tions by principals, or classroom observations, or examination of lesson plans, are examples). This dimension is not addressed in the administration of Title I, and it would require a level of involvement by a federal agency that may not be possible, or may not be judged appropriate.

Some work is being done on the matter of how what the teacher teaches compares with what is on the state test. An ongoing research study in 11 states looks at this question, among others (Anthony Bryk, et al., *The Survey of Enacted Curriculum Project*). In data from eight of the states, correlations were established between the content of instruction as reported by the teachers and the content covered by the state assessment in mathematics in grade 4 and science in grade 8. The findings were that “less than half the intersections of content topics...were in common with the assessment items found on the state...test.” What are we to conclude from the test scores in such situations? Does standards-based reform as contemplated and enforced by Title I include the proposition that students should be held accountable on such tests? And if not, is there really a legislative solution for changing what is happening in these eight states?

F. The Test is Not the Treatment.

I believe that what is good about standards-based reform is what was contemplated at the outset, and exemplified by the experience then available, such as the NCTM standards for mathematics. Content standards were to be a starting point for changing the curriculum and teacher development. A lot of this was actually done in many states — curriculum changes as well as necessary changes in instructional materials and pedagogy were made. In fact, the impact was visible enough that backlashes occurred in some places where there were disagreements over adopted changes. And, appropriately, further

changes were made in the math standards. Those states understood that the test that incorporates the content of the revised curriculum and the instruction that delivers it to the student in the classroom can be an important tool for monitoring the effect on the actual learning of the students, for looking at results, and for taking corrective steps of one kind or another.

As a general matter, I think that professional educators and those whose academic, research, and professional practice involves test theory and construction would be in agreement. Certainly there are many, in what constitutes a political movement to change the schools, who are at least comfortable with the proposition that we can start with constructing a test, give the test, discover that the students do poorly on it, and expect that teachers will change their instruction to enable the students to do better. It could be called test-driven instructional change.

This approach comes very near to what happened in the state of Virginia with the Standards of Learning (SOLs). This is a case where “standards” is used interchangeably with a test, or somehow embodied in the test. In the first test administration, 97 percent of the schools failed, based on the scores that were deemed to be acceptable. Commenting on these dismal results, the chair of the state school board said that when the teachers and students came to learn what was being tested, the students would do better. What is happening now in Virginia is that resources are being identified at the local or state level to institute what amounts to a parallel curriculum. There is a large amount of disarray, many teachers and parents are upset although political support has not eroded. However, the school board did make a number of changes in response to criticism.

An analysis of all that can go wrong in test-driven instructional change, and the inequities and unfairness that can result, would take another paper.

I assume that such test-driven instruction was

not the objective of drafters of the 1994 Title I legislation. I know of no reason to think that it was. But there is a strong current of opinion that believes reform can be achieved by making the test the treatment. So, where a state has approved content standards, performance standards (whatever they are, beyond a cutpoint on a test) and an aligned assessment, all we may really have that is operational is the test. I think such a situation is counter to the goal of raising real student achievement. It will harm the education enterprise, and it will be unfair to students and teachers.



What is advocated here is a reexamination of the provisions of the law dealing with standards-based reform, and a refashioning of the legislative approach to make sure that what the legislation requires is actually resulting in what is intended by it. This reexamination would include looking at what is happening at the district, school, and classroom levels as a result of the way states have (and have not) come into compliance. This will lead to an understanding that the benefits of standards do not flow simply from examining the document labeled “content standards,” or “performance standards,” and from looking at a test to see if the questions somehow reflect these documents.

The problem that must be faced, however, is whether federal law can (or should) prescribe the detailed process to be followed from the creation of such state-level documents to what happens at the level of the classroom, with the approval and follow-up system that would have to be put into place in the Department of Education. Title I dictated a process that had to be statewide, for all students, in order for the states to qualify for substantial sums under the federal law. Considerable consternation would result if the federal government were to extend such statewide control even

deeper into state policy and practice. This paper has focused on the process of education reform, and the steps to be taken in putting standards-based reform into place. There have been other problems and considerations that are important in doing a good job and in gaining public acceptance. There are legitimate criticisms of the quality of the standards in some states, in clear communication to parents and the public of what the state is trying to do and why, and in explaining what the test is measuring and what the results say about student achievement and deficiencies. These are not matters, however, that can be addressed in federal legislation, but need to be addressed if reform is to be sustained and effective.

III. GARNERING SUPPORT FOR CHANGES.

Support for major changes would have to be built by expanding the knowledge base as to what

has actually been happening in many states, and how Title I is involved in the dynamics of implementation of standards-based reform. Creating the knowledge base could come from hearings in the two responsible committees of Congress, and from studies commissioned to inform the committees. It may be the case, however, that the problems I have laid out would not impress the people who constitute the political force for what I consider an over-reliance on standardized testing, in a corruption of the original formulation of standards-based reform. On the other hand, they may not object to a system that returns reform to what I believe was contemplated in the early 1990s, as long as there is continued use of testing for monitoring the results and ensuring accountability.

Paul Barton is a senior associate in the Research Division of Educational Testing Service. The views he expresses are his own, and do not necessarily reflect those of ETS officers or trustees.