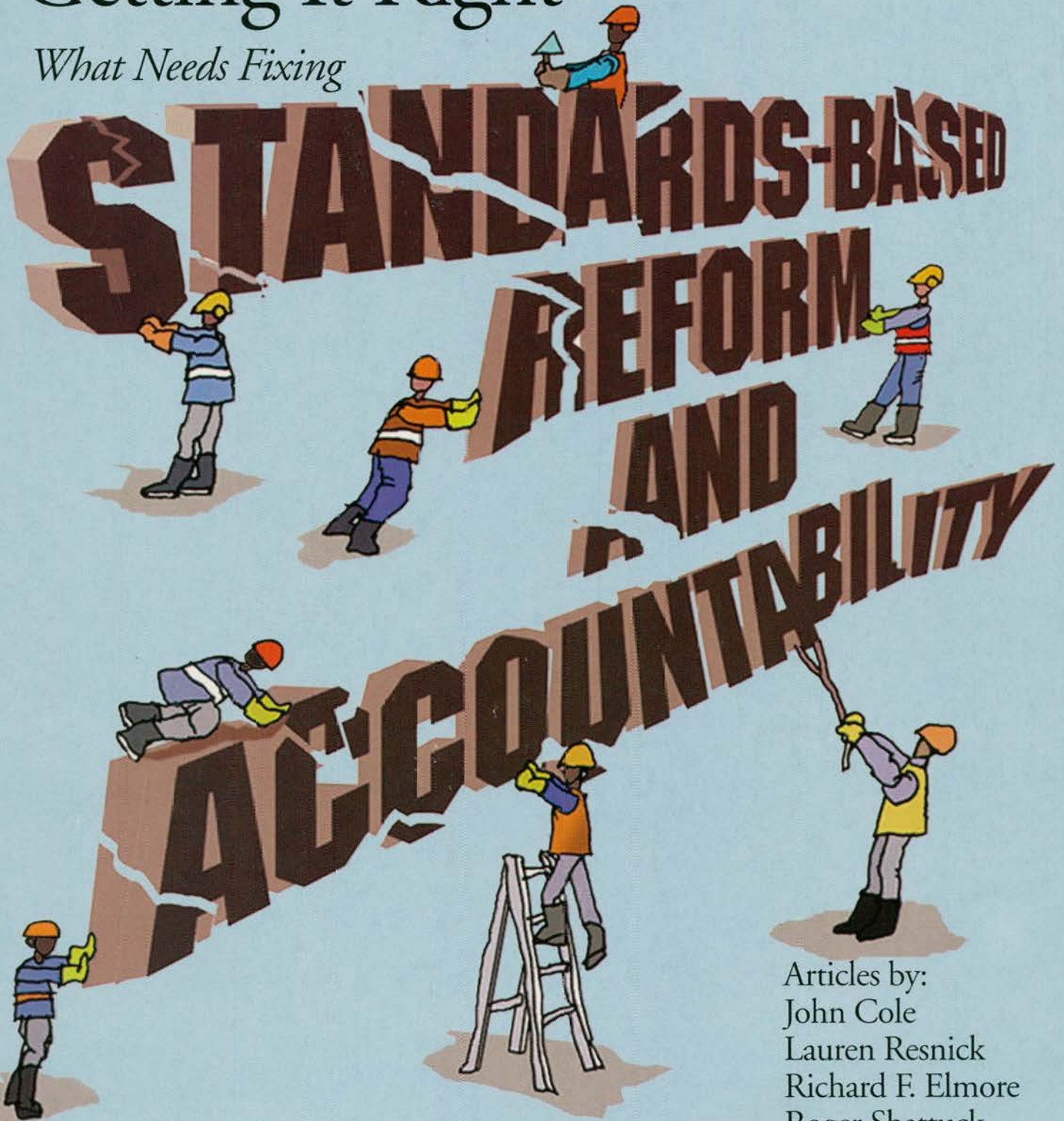


# AMERICAN **Educator**

## Getting It Right

*What Needs Fixing*



Articles by:  
John Cole  
Lauren Resnick  
Richard F. Elmore  
Roger Shattuck  
Nancy Kober



“LET US NOT BE CONTENT TO  
WAIT AND SEE WHAT WILL  
HAPPEN, BUT GIVE US THE  
DETERMINATION TO MAKE  
THE RIGHT THINGS HAPPEN.”

— Horace Mann  
America's first great advocate of  
public education



You know how it feels to see changes that need to be made, to be ready for a new challenge, to be eager to make a difference. It's time to step up! It's time to apply your classroom experience and take your career to the next level.

Presently, there is a shortage of qualified principals, and seasoned teachers—like you—are the best candidates to fill those jobs. Accountability and raising student achievement are in the forefront as never before. With these factors shaping the job, new routes to the principalship are opening every day. If you haven't already considered a career as a middle level or high school principal, now is the time!

NASSP stands ready to prepare you for the challenge. We provide the support, advocacy, professional development, award-winning publications, practical resources, and a peer network you can count on. But making the first move is up to you. To help you get started, we're offering a Six-Month Trial Membership *free of charge!*

Now more than ever, our schools need experienced educators leading them. So take the lead and join NASSP, now.

Take notice. Take stock. Take action.  
Join NASSP. You belong here!

NATIONAL ASSOCIATION  
OF SECONDARY SCHOOL  
**PRINCIPALS**

[www.principals.org](http://www.principals.org)

NX0305AE

**Yes, I want to sign up for a FREE Six-Month Trial Membership!**

Please print.

Name: \_\_\_\_\_

School Name: \_\_\_\_\_

School Address: \_\_\_\_\_

City: \_\_\_\_\_ State: \_\_\_\_\_ ZIP: \_\_\_\_\_

Phone: \_\_\_\_\_ Fax: \_\_\_\_\_

E-Mail: \_\_\_\_\_

Home Address: \_\_\_\_\_

City: \_\_\_\_\_ State: \_\_\_\_\_ ZIP: \_\_\_\_\_

Please choose an edition of *Principal Leadership* magazine:

High School  Middle Level

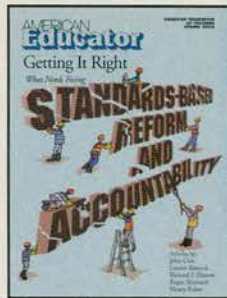
Please send mail to:  Home  School

Please mail completed **NASSP**  
form to: **P.O. Box 3250**  
**Reston, VA 20195-1250**

Associate membership is open to department heads, teachers, graduate students, professors, and all nonadministrators interested in middle level or high school administration. Associate members receive all member benefits, except legal or liability coverage. They cannot vote or hold office. This membership is nontransferable. NASSP membership includes discounted rates on the standard \$85 subscription to NASSP Bulletin, \$30 subscription to NewsLeader, and \$25 subscription to *Principal Leadership*.

For more information, call 800-253-7746 or visit  
[www.principals.org](http://www.principals.org).





Cover illustrated by Bru Associates

## 2 Letters

### 3 Standards-Based Reform and Accountability: Getting It Right

#### 4 Keeping Score

Why Standards and Accountability—Done Right—Are Good for Schools, Teachers, and Kids

By John Cole

When we “keep score,” whether in sports, business, or education, we have to face up to failure—and do what’s necessary to address it. The president of the Texas Federation of Teachers argues that since his state started keeping score by adopting standards and accountability—and also provided the means to help children reach the standards—failure no longer goes quite so unnoticed and unremedied.



#### 20 Building New Knowledge

School Improvement Requires New Knowledge, Not Just Good Will

By Richard F. Elmore

Telling schools to get better is the easy part. The hard part is helping them figure out what the right thing is to do—and how to do it.

#### 28 Curriculum First

A Case History

By Roger Shattuck

A school board member new to the world of K-12 schools discovers to his astonishment that neither his state nor his school district has a curriculum. As luck would have it, he discovers a great curriculum available to all.

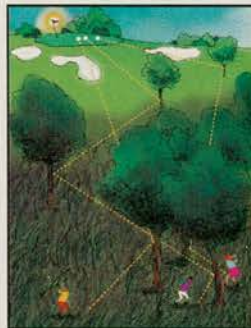


#### 8 Getting Back on Course

Fixing Standards-Based Reform and Accountability

By Lauren Resnick and Chris Zurawsky

Standards-based reform and accountability has brought new attention to teaching our least successful students. But we can’t realize its full benefits or minimize the negative trade-offs without fixing several big problems: inadequate standards; lack of curriculum, instructional programs, and related professional development; poor, narrow tests that are hijacking the reforms; and an accountability formula that identifies not only failing schools, but improving ones too.



#### 35 The “AYP” Blues

Low-Achieving Schools Will Fail—but They’re Not the Only Ones

By Nancy Kober

NCLB’s requirement that schools make adequate yearly progress (AYP) is meant to identify failing schools. But of the six ways that a school could fail to make AYP described by this author, only one is directly related to school quality.

#### 40 Ideas to Consider: How to Make AYP Work Better for Students and Their Schools

### 16 Standards-Based Reform Brings New Attention to Educational Necessities

#### 42 Content Matters

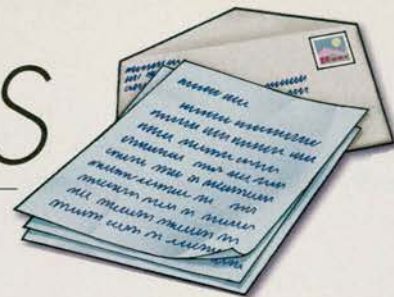
Sometimes Even More than We Think

By Duncan Larcombe

Thanks to her geography teacher’s instruction, a young girl was able to warn many beachgoers that a tsunami was on its way.







The Professional Journal  
of the American  
Federation of Teachers  
Volume 29, No. 1  
Spring 2005

EDWARD J. McELROY  
President  
American Federation of Teachers  
Ruth Wattenberg  
editor  
Lisa Hansel  
assistant editor  
Sandra Hendricks  
copy/production editor  
Jennifer Chang  
production/editorial assistant  
Jennifer Yi  
editorial intern  
Andrew Bornstein  
designer/art director

**American Educator** (USPS 008-462)  
is published quarterly by the  
American Federation of Teachers, AFL-CIO  
555 New Jersey Ave. N.W.  
Washington, DC 20001-2079  
Telephone: 202-879-4420

**American Educator** is mailed to all AFT teacher,  
higher education, and other school-related  
professional members. Non-AFT members may  
subscribe by mailing \$8 per year by check or money  
order to the address above.

Periodicals postage paid at Washington, D.C., and  
additional mailing offices.

**Postmaster:** Send address changes to  
**American Educator**  
555 New Jersey Ave. N.W.  
Washington, DC 20001-2079

**Members:** To change your address, please cut out  
your label with code numbers from the back of the  
magazine and mail it with your new address to the  
address above. In addition, please notify your local  
treasurer.

**American Educator** cannot assume responsibility  
for unsolicited manuscripts. Please allow a minimum  
of four weeks for copyright permission requests.

**Letters to the editor may be sent by regular mail  
to the address above or via e-mail to  
amered@aft.org.**

Signed articles and advertisements do not  
necessarily represent the viewpoints or policies of  
the American Federation of Teachers.

**General advertising office**  
555 New Jersey Ave. N.W.  
Washington, DC 20001-2079  
Telephone: 202-879-4420

**Advertising Sales Representative**  
Karen Dorne  
Karen Dorne Media Sales  
319 Harrison Avenue  
Westfield, NJ 07090  
908-233-6075  
908-233-6081/fax  
kdmedia@webtv.net

**American Educator** is produced with the assistance  
of members of Local 2, Office and Professional  
Employees International Union, AFL-CIO, and  
members of AFT Staff Union. Composition and  
printing are done in 100 percent union shops.  
© American Federation of Teachers, 2005.



## Problem Novels: Too much, too often, too early, or just right?

"Reflections on the 'Problem Novel'"  
by Barbara Feinberg (Winter 2004/  
2005 *American Educator*) is an impor-  
tant contribution to our thinking about  
how we are helping/hindering our  
young people grow up.

I grew up on books and films that  
inspired—that built courage in the face  
of disappointment and sorrow, such as  
*A Tree Grows in Brooklyn*. And the  
films—oh, those films: *Madame Curie*,  
*National Velvet*, *The Yearling*, *Young  
Tom Edison*. I know that I date myself,  
but the books and films I saw nurtured  
me as a person. What worries me, as it  
does Barbara Feinberg (who did a great  
job on your article), is that our youth  
are being fed a diet that does not nour-  
ish, but instead, destroys and disrupts.

I am not for giving our kids pabulum,  
but there has got to be a stronger mid-  
dle ground. As a grandparent of two  
boys, I want to know the titles that ex-  
plore real life and yet strengthen abili-  
ties to grow and survive. **I ask your  
readers who know some of  
these books to contact me at:  
dorothyrich@starpower.net.**

Again, thank you for the article and  
may every librarian read it.

—DOROTHY RICH

President, Home and School Institute  
Author, *MegaSkills Programs*  
Washington, D.C.

Thank you for publishing Barbara  
Feinberg's article exposing the problem  
with the "problem novels" that our  
middle school students are required and  
encouraged to read. I feel comforted to  
know that my perceptions are shared  
and that my feeling of responsibility for  
the messages we feed children through  
literature is also shared. Feinberg has  
eloquently reminded us that children,  
and perhaps the rest of us, thrive on the  
concept of "open destiny" or hope, and

that children live in a world that com-  
bines reality and fantasy. Perhaps kids  
should not be required to consume  
these books as a steady diet, but those  
who do find solace in them should have  
them available.

—LYNN ROBINSON

As I read Barbara Feinberg's essay on  
"problem novels," I became increasingly  
exasperated with her solemn, aggrieved,  
self-righteous tone. The kinds of books  
she would have children read are the  
kind I would do anything to avoid.

Because I was allowed to read any-  
thing I chose when I was growing up, I  
came to understand that catastrophe  
and suffering were a part of life. My  
parents never imagined that I was so  
vulnerable that I needed to be protected  
from something—anything—that  
might happen in books. Writers write  
(and readers read) about calamity be-  
cause it is interesting. My parents un-  
derstood this, and they understood that  
I understood it.

I think Feinberg misunderstands the  
whole point of fiction. She seems to  
think novels should provide their read-  
ers with chirpy, perky little exhorta-  
tions—like "Be brave!" What a reader  
should ask of a novel, however, is in-  
finitely more subtle: Whether it is  
beautiful, and whether (yes, even a  
child can answer this) it is true.

—ALEX JOSEPH

The Fashion Institute of Technology  
New York, N.Y.

## Correction

In the Winter 2004-05 issue, *Ameri-  
can Educator* incorrectly identified the  
Web address for Beacon Press. To  
order a copy of Barbara Feinberg's  
book, *Welcome to Lizard Motel:  
Children, Stories, and the Mystery of  
Making Things Up*, please visit **www.  
beacon.org**, and enter the promo-  
tional code FEINBERG for a 20 per-  
cent discount. We apologize for any  
inconvenience this caused our readers.



# Getting It Right

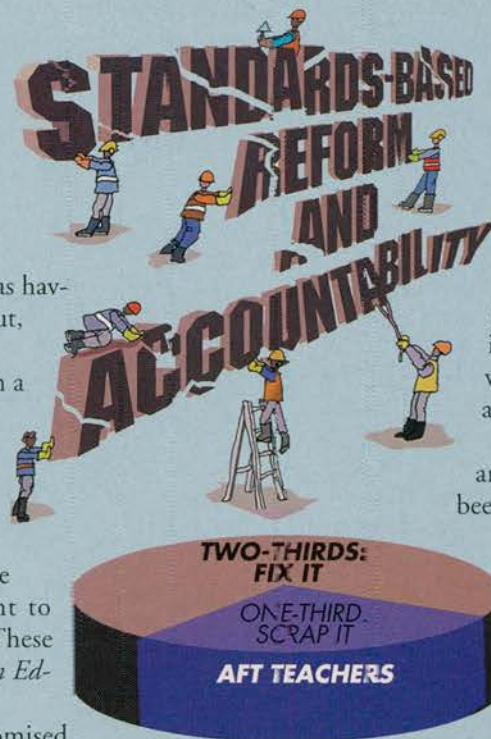
## What Needs Fixing

In a survey of AFT teachers last year, two-thirds said the No Child Left Behind law was having a negative effect on public education. But, by the same two-thirds margin, they said they wanted the law fixed, not scrapped. Likewise, in a 2002 poll, two-thirds of AFT teachers said standards-based reform and accountability is the “right approach for improving education, but there must be improvements in the way it is carried out.”

What needs fixing? And why, despite the problems, is there a continued commitment to standards-based reform and accountability? These questions are the focus of this issue of *American Educator*.

The idea of standards-based education promised high educational standards and a common, equitable curriculum for all kids; tests that measured progress toward the standards; special attention for children struggling to reach the standards; instructional materials and professional development based on the curriculum (bringing the quality of our educational system in line with those of other high-achieving countries); and an accountability system that targeted resources and attention where they were most needed. As Texas Federation of Teachers President John Cole notes in our first article, school systems unhampered by public standards and accountability can—like other institutions—act neglectfully, especially toward the least advantaged.

But, for reasons explained by authors Lauren Resnick and Chris Zurawsky, inadequate tests, and accountability based on them, have often gotten dangerously out in front of the other elements of standards-based reform, threatening the very educational quality we’re trying to build; the authors outline the attention that must be paid to the lagging pieces. As Roger Shattuck observes in his article, many communities still don’t have curricula worthy of the name. And, as Richard Elmore notes, most school districts still barely understand, much less have addressed, the huge challenge of building faculty and school capacity to dramatically lift student achievement. We have no systematic way to make sure that what’s known about good instruction gets to all teachers—and no systematic way to learn what we don’t yet



know. As for accountability, Nancy Kober explains why the adequate yearly progress formula in NCLB can (and increasingly will) identify the wrong schools as failing.

But for all the undone work and broken promises, there have been notable successes (described in “Getting Back on Course,” p.8, and box, p.16). And the new visibility of test results, which make brilliantly clear just how far behind our poorest students are, may be finally galvanizing a new understanding of just how much

harder it is to bring children from poor schools to high proficiency levels—and, therefore, how much better, and better supported, these schools need to be than other schools.

Case in point: States, under their own constitutions, are typically responsible for providing their children with an adequate education—but time and again, they’ve battled in courts to define “adequate education” at a low level. Now, public test results showing low achievement, especially among poor students, are forcing states to grapple more seriously with what it takes to offer a decent education. The National Council of State Legislatures’ recent report on No Child Left Behind acknowledged that to meet the goal of the law—to bring poor children in particular to the required achievement levels—substantial new investments will be necessary to, among other things, build teacher capacity, intervene with struggling students, and increase access to early childhood education.

The understanding that educating our poorest, furthest behind students will require enormous work and resources has been hard to come by. But it may be spreading, largely thanks to the standards and accountability movement. AFT teachers are right: There’s lots of fixing to do to get it right, but standards-based education is worth the effort.

—EDITORS



# Keeping Score

## *Why Standards and Accountability—Done Right— Are Good for Schools, Teachers, and Kids*

By John Cole

Back in the 1980's, when Texas education reform got underway, I was often asked: "Why is the teachers union supporting the new school reform law?" The 1984 law required a lot of new accountability for teachers and students, including district report cards that contained information on student test scores and a high school exit exam. I'd be asked, "Isn't it a lot easier for you all when people aren't breathing down your neck about test scores?"

A lot of teachers were also dubious of the law. They worried that it would force them into a teaching straitjacket and that it might mean that a lot of decent kids would fail classes and might not even get a high school diploma. But I have always argued that standards and accountability, combined with the support that teachers and kids would need to reach the standards, are good for public schools, teachers, and kids—especially poor kids.

Why do I think this? Let me start with two anecdotes. First: I am an avid Dallas Cowboys fan. Back in 1993, my favorite running back, Emmitt Smith, was requesting something like \$13 million for a four-year contract with the Cowboys. I immediately wrote to the Cowboys and offered to serve as their running back for much less—perhaps one-tenth of that amount, even a hundredth of that amount. Amazingly, the Cowboys never responded to my generous offer. I asked myself, "Why—why do they want to pay this guy \$13 million when they could get me for just \$100,000?" The answer, I think, has to do with keeping score. If you don't keep score, the quality of your players really doesn't matter. In Texas football, we keep score. The Cowboys keep score. And they care about scoring well. That's obviously why they're willing to pay Emmitt Smith all that money even though they could have me practically for free.

Here's the second anecdote: To pay for college, I sold in-

---

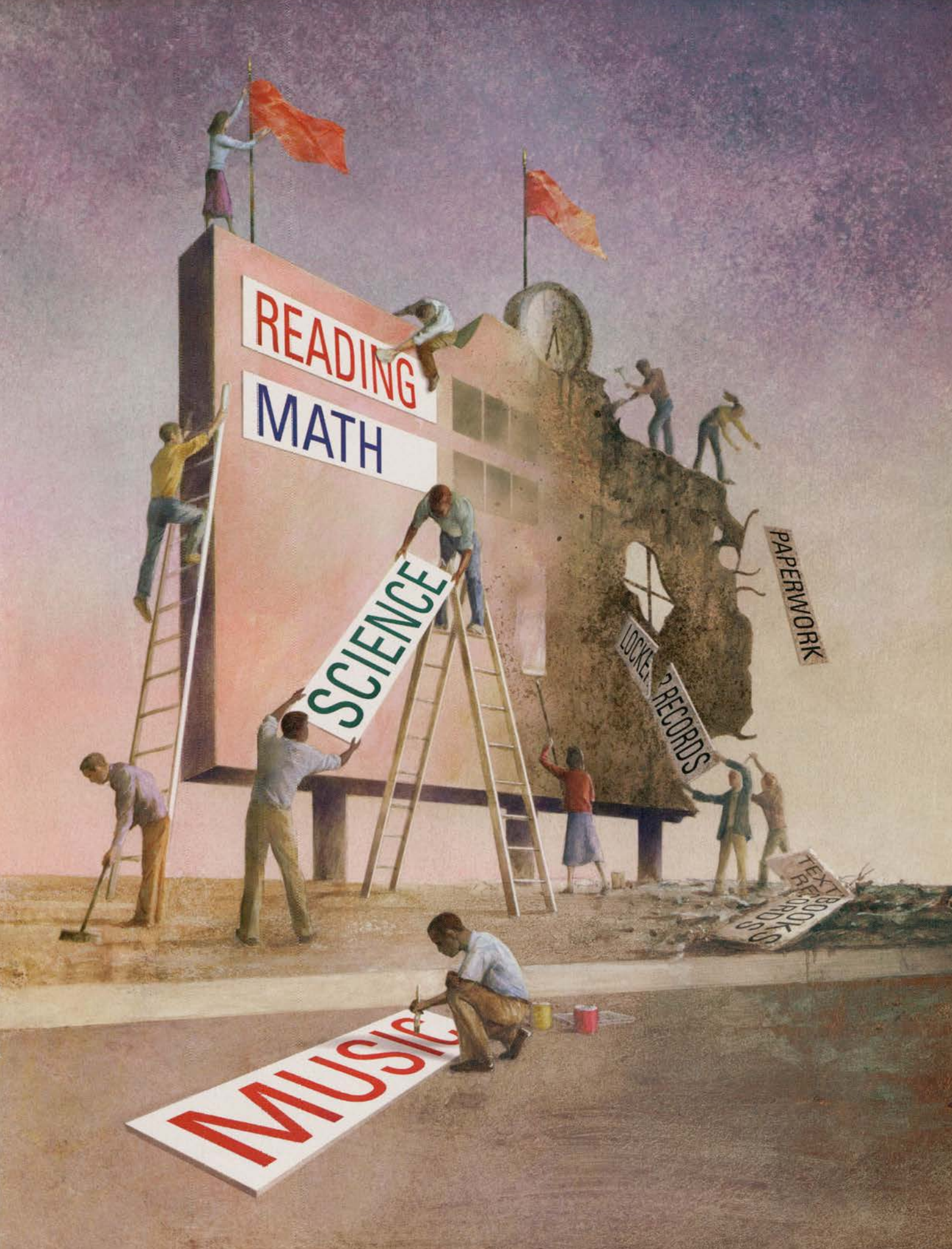
*John Cole is president of the Texas Federation of Teachers and a vice president of the AFT. Previously, he was a teacher in Corpus Christi, Texas, and the founding member and president of the Corpus Christi Federation of Teachers.*

urance for a while. In my office there were about four guys and a manager. Three of us were young kids like myself and one was this older gent who had been selling insurance forever. This guy *never* came to the office. He missed *every* staff meeting. His accounts often didn't balance, which would have been bad news for the rest of us, but not for him. If his account was \$10 over, he took \$10 out. If it was \$10 short, he pitched \$10 in. The manager treated him like he was some sort of deity. The rest of us were treated more or less in accordance with our just desserts. Why was this? Because in the insurance business there's a way of keeping score. That manager's salary was determined by the amount of insurance sold out of our office. That older gent sold more than the rest of us put together—probably twice as much. So, the manager didn't care about whether he came to staff meetings or even behaved rudely (which he often did). What the manager knew was that this was the guy who produced good paychecks. Whatever would keep that guy selling was important to the manager, the rest was trivial.

I am a product of several decades of Texas education. I actually went through the schools here and then began teaching. And I can tell you, during all that time, *no one* was keeping score—or, to be more precise, no one was keeping score about matters like student achievement. And, when a school system doesn't keep score on student learning, there's not a lot of pressure for learning to improve. That means there's not a lot of pressure to pay the kinds of salaries that would attract qualified teachers. It means there's not a lot of pressure to make sure poor schools have books that aren't torn and old as dirt. It means martinet principals can focus on trivial matters like locker records instead of results.

Let's start with my initial years as a teacher in Corpus Christi. Schools in Corpus Christi weren't desegregated until 1976. So when I started teaching in the late 1960s, we had three sets of schools—one for whites, one for blacks, and one for Hispanics. I taught in the Hispanic junior high school. We were blessed, I suppose, in that we got the text-





READING  
MATH

SCIENCE

LOCKER RECORDS

PAPERWORK

MUSIC

TEXT BOOKS



books right after the white junior high school was finished with them—when we finished with them, we sent them over to the black junior high school.

The school had a lot of dedicated teachers, but as an institution, the public school system didn't really care too much about what went on in the school that I taught in—or in the other schools that Hispanic and African-American kids attended. For example, there were no standards for coursework. We had valedictorians from some of these schools who couldn't get into college because they hadn't taken the right courses. There were many places where kids took the same remedial math course four years in a row under a different name. They never got to algebra, never got beyond arithmetic. When I started teaching, I was told that four percent of the kids at my junior high went on to graduate from the Hispanic high school in Corpus Christi; the other 96 percent dropped out along the way. As far as I could tell, not one person cared if I ever taught a lick.

Here's how my school worked: One of the teachers was absolutely beloved by the principal. He was a coach who was assigned to teach English. He got every 16mm film that he could order and he showed one every day until the last couple of weeks of school when the film library was closed. Then, to wrap up the school year, he bought a bunch of coloring books. The grades in his class were based primarily upon attendance and comportment.

But this teacher never got into trouble for his behavior. (In fact, later on he became an administrator!) Again, I asked myself why? But the answer was easy. What, after all, was important to this principal? Certainly not learning. Above all else, what was important to him was that nobody showed up at his office door. He didn't want to see angry parents or kids complaining.

In this principal's mind, I was a terrible teacher. I complained that we had no program for the kids who didn't speak English. I complained that we were short of textbooks and that the ones I had were missing pages. I complained that we needed to get some eyeglasses for the kids whose parents were too poor to buy them. I was a source of problems and disruption; I caused grief for that principal.

The coach-turned-teacher, on the other hand, was a model that everyone was supposed to look to and admire. Why was that? Well, nobody kept score of the students' learning. The school system did keep score of some other things, though. If a teacher's textbook records showed up in disarray, that was a problem. If a teacher's locker records were in disarray, that was big trouble. But during the entire time that I taught, I never once had anybody ask me about the students' learning.

**I**n the late 1980s, several years after I became president of the Texas Federation of Teachers, I served for two years on an official state committee charged with recommending what indicators of performance should be included on school report cards. Our hope was that they would include information on test scores, dropout rates, and other factors. Part of my job was to hold public hearings in different parts of the state. The only people who came were school board

---

These school board members and superintendents knew good and well where education was happening and where it wasn't, but clearly they didn't want the public to know. Pressure would build to improve those schools.

---

members and superintendents, by and large. And at each hearing it was the same. I could have written the script. I would go through my presentation and have my charts. And they'd have one question: "Are we going to publish this information?" Well, yes. I'd tell them that the idea was to make this report card available to parents and the public. After a moment of general consternation, there would be an observation: "Wait a minute, if we do this, nobody is going to want his kid to go to this school over here." And then somebody else would say, "Oh, and what about that school over there? Everybody will want to be in that school."

"Wait a minute," I'd say. "Are you telling me there's a school in your district right now that doesn't teach kids, you know of it, and you're not telling anybody—you're just letting it sit there?" These school board members and superintendents knew good and well where education was happening and where it wasn't, but clearly they didn't want the public to know. Pressure would build to improve those schools. They'd actually have to find resources for those schools, offer salaries that would attract qualified teachers, and get them textbooks that weren't ripped up and old. They'd have to make sure kids were learning something before they were promoted or given a high school diploma. It was a lot easier for them to just pretend there was no problem. It was a conspiracy of silence. And, there was no way to blow the whistle on it because there was no objective way to compare student achievement across schools and districts.

## Standards and Accountability Blow the Whistle

In Texas, we started keeping score when the school reform law passed in 1984—long before George Bush was governor, I should point out. And, because we started keeping score, that marked the beginning of the end of the conspiracy of silence. Yearly testing in reading, writing, and math in grades 1, 3, 5, 7, 9 and 11 began right away; and the 11th-grade test became a requirement for graduation in 1987. Not surprisingly, many districts resisted the idea of a state exam; they each wanted to decide on their own test, their own passing score. Districts had long played a game in which they would give their own test, and if scores were good, they'd use them to say how great their schools were;



---

Now, is that all we did—put tests and accountability into place? Absolutely not. Tests don't teach and tests don't produce miracles. We put the test in place, we put the standards in place, but we also put tons of new money in place.

---

they ignored the scores if they were low. We ended that game once there was a single state test.<sup>1</sup>

Keeping score has made a world of difference—it has ended that conspiracy of silence, or at least made it a much harder game to play. There is absolutely no question that we've ratcheted up the quality of education in Texas dramatically. The test we gave 6th-graders this year was harder than the one we gave 11th-graders back in 1987. And despite dire predictions to the contrary, while we've raised the difficulty of our standards and curriculum pretty steadily since then, the drop-out rate has remained pretty constant (pretty constantly awful, I should say). But we have roughly the same percentage of kids in school, and they're passing tougher tests at higher rates. We haven't shut down the achievement gap between white and other children, but it's diminished.

Now, is that all we did—put tests and accountability into place? Absolutely not. Tests don't teach and tests don't produce miracles. We put the test in place, we put the standards in place, but we also put tons of new money in place. The standards and accountability have to be there, otherwise the districts get money, and who knows where it goes? It often winds up paying for wonderful lessons on self-esteem (or worse), but not the things that effect academic achievement.

But you can't expect to raise standards and get better teaching unless you commit the resources to pay for the good salaries that will lure qualified teachers into the classroom; to pay for the professional development that teachers need to teach better; to get extra help to the kids who really need it and the schools that really need it. When we passed the 1984 reform, we added 13 percent to our state aid per pupil. And that wasn't the end of it. We kept pumping in new money so that between 1984 and 2000, state aid per pupil increased by 24 percent (in constant dollars). And we didn't just add new money—we redirected the state's resources so that low-wealth school districts and school districts with high concentrations of disadvantaged children received the bulk of it. It was a revolution. We also created a minimum standard for teachers; it was suddenly much harder for administrators to hire unqualified people and call them teachers. Starting in 1986, all teachers had to take a basic reading and writing test; if they couldn't pass it, they lost their teaching certificate. But we also increased salaries, spectacularly so in the poorest districts, so that when new

teachers were hired, we were able to attract teachers who met the higher standard.

In 1999 Texas enacted legislation, which the Texas Federation of Teachers initiated, that made passing the third-grade reading test a requirement for promotion to fourth grade. The requirement kicked in as of 2003 because that's when the 1999-2000 crop of kindergartners reached third grade. But that legislation didn't just create a barrier to promotion for those kids, it provided resources to pull together the key ingredients for success, including professional development for their teachers, diagnostic assessments, and immediate interventions. Beginning in 1999 with kindergarten teachers, and adding a grade each year, Texas provided paid professional development opportunities to virtually all the state's K-3 teachers. By 2001, nearly 60,000 teachers had already received the training. The student failure rate on this third-grade reading test prior to 2003 (the year it became a requirement for promotion) was about 20 percent. With professional development, early assessment, interventions, and accountability, we cut that failure down to about four percent in 2003.<sup>2</sup>

**T**ests don't teach; accountability on its own doesn't make teachers teach better. Shutting down schools when you have no better strategy for making them work the second time around does no one any good. But accountability makes people keep score. It helps stop the conspiracy of silence. And that helps get the resources flowing to schools—and it helps to make sure the resources are used well. It helps people see that giving out high school diplomas doesn't mean you've educated the kids. And, as with running backs, it helps people see that just calling someone a teacher doesn't make it so.

Clearly there's still much to do to increase achievement in Texas. From the 1980s to today, one of the main questions has been how to increase the level of difficulty on the student tests and provide the support that teachers need to make sure that students can pass. We still do not have the grade level or graduation tests where we want them to be. Salaries are higher, but still not where they should be. And, after years of support, this year the Texas legislature seems bent on grossly underfunding education. But we've come a long way in the last 21 years, and it would not have occurred without standards, professional development, additional resources, and the accountability that comes from the test. □

## Endnotes

<sup>1</sup> Based on my experiences in Texas, I believe that No Child Left Behind (the federal legislation mandating school improvement) gave away the farm by allowing all states to have their own standards and tests. Without a common standard and a common test, there's a strong incentive for individual states to lower their passing bars thus making it look like their students are highly proficient.

<sup>2</sup> This year we ratcheted up the standards again, so the passing rate on the first administration was 89 percent. I expect the final passing rate will be higher, but not as high as it has been for the past two years. In the past, whenever we ratcheted up standards, more money flowed to school districts to help them meet those standards. But this year the governor and legislature have cut money from education. So far, they have eliminated our master reading and math teachers, as well as our remedial programs.



# Getting Back on Course

By Lauren Resnick and Chris Zurawsky

**T**he last 15 years have witnessed a profound sea-change in American education. Labeled “standards-based education,” the shift has involved important changes in the basic mode of operation of our schools and has greatly affected the lives of teachers and other educators. It has entailed a greater emphasis on academic achievement, a more urgent commitment to equity in academic opportunity (especially for minority and other academically at-risk students), a shift in the locus of decision-making about what should be taught to students—away from individual teachers and local schools toward districts, states, and even national standard-setting bodies—and much greater accountability, meaning consequences for students and/or schools when academic goals are not met.

Taken together, these changes are creating difficult challenges for front-line educators. Educators have been asked to teach all students to high levels (levels once reserved for the best prepared and most privileged students)—but because of the widespread lack of adequate support and preparation, teachers frequently feel they are being told to do the impossible. States and districts are telling schools and teachers what they should teach and how they should teach it, at levels of detail rarely experienced in the recent history of Amer-

ican schooling. The push for better performance in the core subjects of math and reading often seems to be driving nearly everything else out of the curriculum. In a recent survey by the Center on Education Policy, “27 percent of districts reported that time devoted to social studies had been reduced, almost a fourth reported that time in science, art, and music had been reduced, and 10 percent reported that time given to physical education had been reduced” (CEP, 2005). For many people, it seems as though prepping for tests is taking up more and more of the school day (Olson, 2002), and there is little time left for deep reading, extended essays, science experiments, or theater productions. In some localities, parents have protested, school boards have resisted, and even several state legislatures have called for roll-backs in the federal No Child Left Behind Act, which the legislatures believe is forcing federal control and “standardization” upon a land proud of local educational independence.

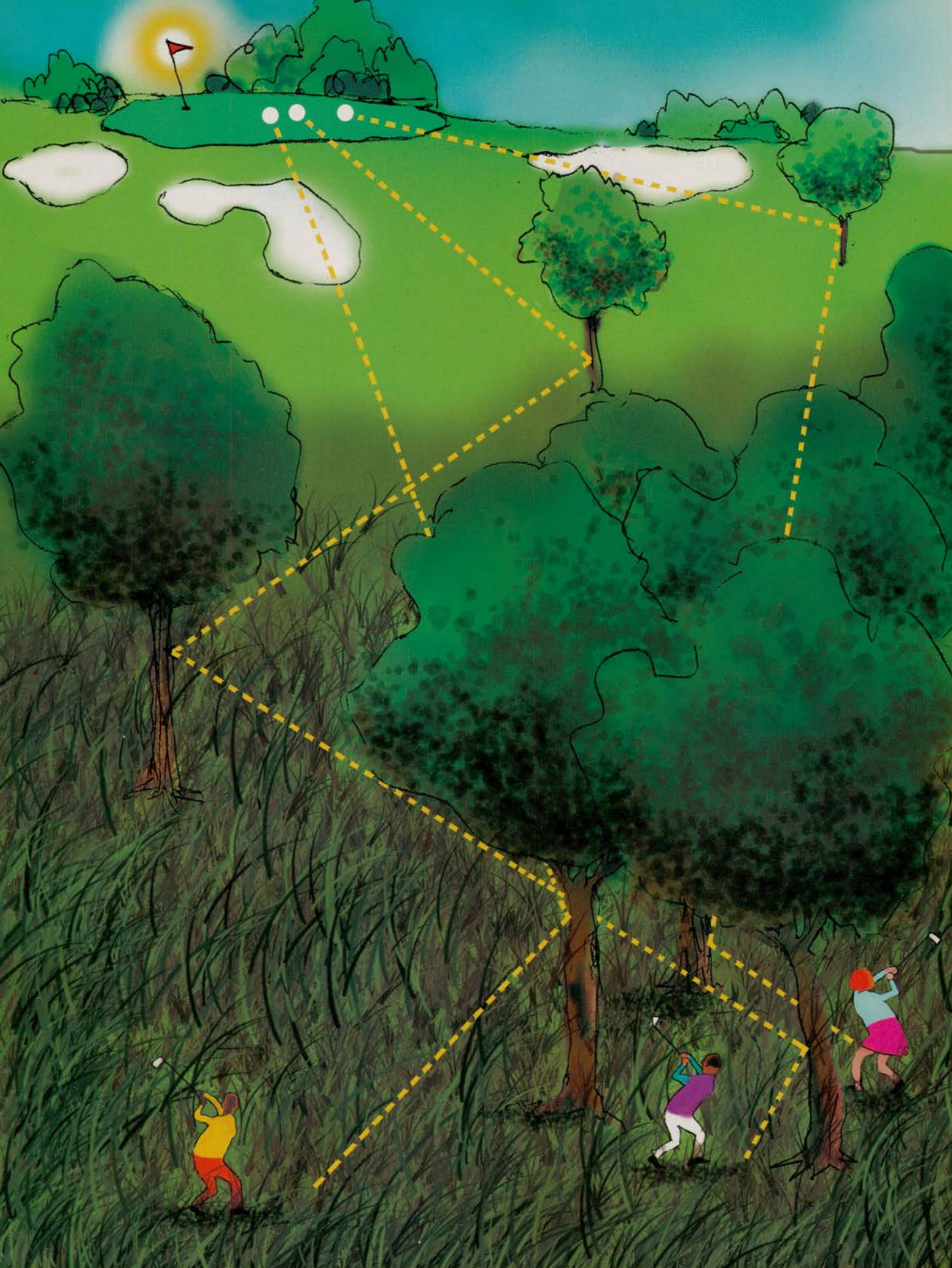
**A**t about 15 years of age, the standards movement is in its adolescence, and many are already preparing to kick it out of the house. Before we give up on our unruly teen, however, let’s take a clear look at what we have to be proud of, what flaws we need to address, and what might be the benefits of pressing ahead. We have serious questions to ask: Where did the idea of standards as a foundation for an education system come from? And how did tests come to run the show? Is there any evidence that poor and minority students are benefiting from a standards-based system? Is overall academic performance really improving? Or, are we busy tearing down an education system that was pretty good and pretty equitable? In short, is there enough gain to warrant the pain?

Our own answer is a qualified yes. We think that the effort to create a standards-based system for American schools is just and relevant, and it is starting to work, especially for the poorest children in the most challenged schools. For the first time in our history, American schools are truly focused on fostering the academic achievement of all students. And

---

*Lauren Resnick is director of the Learning Research and Development Center (LRDC) at the University of Pittsburgh, and founder and director of the Institute for Learning, which provides professional development to urban school districts. She is also editor of Research Points, a publication of the American Educational Research Association, and is co-founder and co-director of the New Standards Project, which has developed educational standards and assessments for states and school districts. Chris Zurawsky is LRDC’s communications director and managing editor and an issue writer for Research Points. The authors wish to thank the National Science Foundation for partial support of the preparation of this paper. The opinions expressed in this paper are the authors’ and do not necessarily reflect those of the Foundation.*







it is happening at the same time that we are devoting unprecedented attention and care to the education of children who come from low-income, minority, and immigrant families. We can see this in a decade's worth of increased budgets for early education, of increased state and federal budget allocations for K-12 education, and of what appears to be a growing commitment at the state and local level to supporting programs aimed at helping the lowest performing schools and students (CEP, 2005). For example, state funding for prekindergarten (for which most states limit eligibility to low-income and other at-risk children) increased from about \$267 million in 1988 to \$2.54 billion (in constant dollars) in 2002-03 (Barnett, 2005; Barnett et al., 2004); federal funding for K-12 education has increased from \$29.6 billion to \$59.7 billion in constant dollars between 1990 and 2003, though the increases have now slowed (Sonnenberg, 2004); and, on average, states' real per capita expenditures on elementary and secondary education increased by 24 percent between 1988 and 1997 (Merriman, 2000).

The full picture of student achievement growth over the past decade can't yet be drawn. Much has happened that will never be captured in data, much data linger unanalyzed, and, not surprisingly, much data remain in dispute. Thus, the debates about how the positive and the negative effects of standards-based reform balance out will continue. But for us, the weight of the evidence indicates that student achievement, especially among the most disadvantaged students in the poorest districts, is increasing—and is doing so thanks in large part to the reforms and resources generated by the standards-based education and accountability movement.

As part of our work at the University of Pittsburgh's Institute for Learning, we regularly examine student achievement data from our partner school districts. In these districts, where standards are being translated into systematic programs of instruction and are increasingly backed by professional development, the effects are now clearly visible in elementary school reading and mathematics performance. To take three examples: Since 1999, the Saint Paul Public Schools have made significant progress in raising academic achievement in reading and math, especially among minorities. Between 1999 and 2004, the percentage of 5th-grade students who scored proficient or above in reading on the Minnesota Comprehensive Assessments went from 31 percent to 54 percent for American Indian students, 29 percent to 49 percent for Hispanics, and 26 percent to 46 percent for African Americans. In Austin, Texas, every student group showed significant gains in passing the state reading assessments for 3rd and 5th grades between 2003 and 2005. The passing rate for African-American 3rd-graders, for example, grew from 64 percent to 78 percent; and for 5th-graders, it grew from 49 percent to 60 percent. There was slightly smaller but still significant improvement for Hispanic and economically disadvantaged students. Providence, R.I., is also showing gains in student achievement. In 2002, only a handful of schools met NCLB's target, but in 2004, almost all schools met the target for all ethnic groups.

These results don't appear to be isolated. According to the Council of Great City Schools (CGCS, 2005), "55.3 per-

cent of 4th-grade students in the Great City Schools scored at or above proficiency levels in math in 2004, compared with 50.8 percent in 2003 and 44.1 percent in 2002." Results in reading are similar, with proficiency rising from 43.1 percent of 4th-grade students in 2002 to 51.0 percent in 2004. Perhaps the most encouraging data come from the National Assessment of Educational Progress, which showed large gains in math and smaller but promising gains in reading during the 1990s (Jennings and Hamilton, 2004).

**W**e need to celebrate all of these gains. But to say this is not to say all is well. The achievement goals of this education reform movement are ambitious and the large-scale efforts to reach them are recent and previously untried. As a result, the on-the-ground path to improvement is largely new and uncharted and filled with all the extra work and frustration of trial and error—false starts, wrong paths taken, constant rethinking. Further, the gains to date appear mainly to have elicited a rise in the achievement floor.

If we are to expand on these gains, we must figure out how to amend and facilitate and thereby strengthen our national experiment in school reform. To do so, we must first go back in time and consider the conditions that launched this movement and gave rise to the high hopes for standards-based education reform.

## I. The World that Launched Standards-Based Reform

Put yourself back in the Zeitgeist of roughly 1980 to the mid-1990s. Our nation's schools had been expanding access in previously unimaginable ways. With *Brown v. Board of Education*, the Supreme Court ended *de jure* segregation, thereby requiring the previously all-white school system to address the needs of black students, a challenge it was still working to meet 30 years later. The Education for All Handicapped Children Act (now known as IDEA), passed in 1975, guaranteed a free and appropriate education to children with disabilities—from the learning disabled, to the blind, the emotionally disturbed, and the mentally retarded. At the same time, new waves of immigration, mainly from poor countries in Latin American and Asia, had increased the number of students who spoke English with difficulty from 2.2 million in 1979 to 4.2 million in 1995 (Mandlawitz, 2005). By the end of the 1980s, America's public schools were serving all of these children, children who earlier in our history were segregated, isolated at home, or sent into the workforce at an early age. With all of this, the percentage of high school graduates (as a ratio of the 17-year-old population) increased from 51 percent in 1940 to 74 percent in 1990 (National Center for Education Statistics, 2003).<sup>7</sup>

America could be proud that so many young people had access to an education. But what was the quality of the education they had access to? A crisis was first signaled publicly

<sup>7</sup> Since researchers have yet to agree on the proper way to calculate graduation rates (e.g., whether or not to include people who have earned a GED), readers have probably seen higher and lower graduation rates than these from NCES.



The on-the-ground path to improvement is largely new and uncharted and filled with all the extra work and frustration of trial and error.



in the U.S. Department of Education's seminal 1983 report, *A Nation At Risk* (National Commission on Excellence in Education, 1983). Using data from international comparisons of educational achievement and research on course-taking in American high schools, it concluded that the U.S. was at risk of losing its lead in mathematics, science, and technology. Among its findings:

- Only 31 percent of recent high school graduates completed intermediate algebra; only 16 percent geography; and, partly because it wasn't even offered in 40 percent of schools, only 6 percent completed calculus.
- For students in the general track, 25 percent of their credits were earned outside regular academic courses, including in physical and health education, but also in remedial English and math and "personal service and development courses, such as training for adulthood and marriage."
- International comparisons of student achievement, completed a decade earlier, showed that "on 19 academic tests, American students were never first or second and, in comparison with other industrialized nations, were last seven times."
- High school graduates weren't cutting it in college. Remedial mathematics courses in public 4-year colleges increased by 72 percent between 1975 and 1980; by the early 1980s they made up 25 percent of all mathematics courses taught in those institutions.

These findings were bolstered by widely-read books on America's high schools. In *Horace's Compromise*, TheodoreSizer (1984) wrote that high school students and their teachers typically had a "bargain" in which the teachers wouldn't demand much effort and in return the kids would be "friendly and orderly." (A similar report came from Ernest

Boyer's [1983] Carnegie Foundation study of high schools.) And, *The Shopping Mall High School: Winners and Losers in the Educational Marketplace* (Powell, et al., 1985) decried the "smorgasbord" curriculum, in which students could load up with remedial classes and courses with such easy-to-mock names as "Applied Communication," "Business Arithmetic," and "Foods" and never take a difficult math course or write a research paper—and still graduate with a high school diploma.

Throughout the 1980s, the call for higher achievement grew beyond the federal government and academia, spurred by the changing economy. In the early 1980s, the country was struggling against a recession and unemployment that went as high as 9.7 percent (Bureau of Labor Statistics, 2004). Powerhouse companies in Japan and Europe were competing successfully with American companies and, it seemed, jeopardizing our premier role in the world economy. Traditional well-paying jobs were disappearing and many people came to believe that a high-wage economy required a focus on "working smart"—that is, shifting away from jobs in which a strong back and willingness to work were all that was needed to make a good start in America. Not surprisingly, the weight of the business community got behind major education reform.

Along with the push for global competitiveness, increasing attention was being paid to educational equity. The huge achievement gap between black and white was becoming increasingly obvious. As one example, in October 1977, when Florida sophomores faced a functional literacy test that was a new requirement for a diploma, 78 percent of black students—but only 25 percent of white students—failed (Debra P. et al., 1979). And Florida was not alone. According to the National Assessment for Educational Progress (NAEP), throughout the 1980s, black 12th-grade students' scores in reading and math were about equal to those of white 8th-graders—and Hispanic students were not faring much better (NCES, 2000).

The growing perception among employers and higher education professors that the high school diploma had lost its luster; the nervousness about what all this would mean to our ability to compete in the increasingly global economy; the dramatic achievement gaps—all of these contributed to the growing belief among governors, policymakers, business leaders, and Americans generally that something had to be done to dramatically lift the quality of American education. By the end of the 1980s, many researchers and policymakers were beginning to converge on a solution.

### The Promise of Standards-Based Reform

In the 1980s and early 1990s, while dissatisfaction was continuing to build, some policymakers and researchers looked overseas at the education systems that had performed well on a variety of international assessments (Resnick and Resnick, 1985). Virtually all had education systems that were anchored by a national or nationally coordinated curriculum, which outlined in some detail the content and skills that students were expected to learn. Typically, students across these countries studied a common curriculum through at least 4th grade (Germany) and often through 8th



# An American Revolution: A Common Curriculum

*Under Albert Shanker, president of the American Federation of Teachers from 1974 to 1997, AFT was one of the earliest advocates for high-quality, rigorous academic standards. In his weekly New York Times column, Shanker often explained the benefits of getting the standards and accountability movement right. This excerpt from his February 24, 1991, column highlights the benefits of a common curriculum.*

—EDITORS

## By Albert Shanker

In most countries with a common curriculum, linkage of curriculum, assessment, and teacher education is tight. Once you have a curriculum on which everyone agrees, you have an answer to the question of how to train teachers. They have to be able to teach the common curriculum. And you have an answer to the question about the level of understanding and skill student assessments should call for because you can base assessments on the common curriculum.

In the U.S., we have no such agreement about curriculum—and there is little connection between what students are supposed to learn, the knowledge on which they are assessed, and what we expect our teachers to know.

Each of our 15,000 school districts and 50 states has some rights in establishing curriculum. (And this is a nation where people move more often than in any other country in the world.)

In most countries with a national curriculum, tests usually consist of writing essays or solving problems based on what the students are supposed to know. And when youngsters, with the help of their teachers, prepare for these tests by answering questions that were on previous tests, it's a worthwhile educational experience. Writing an essay on the causes of World War I or presenting the arguments for and against imperialism is a good exercise in learning substance and in learning how to organize your thoughts. And the quality of the essay really shows how well the student has mastered the material.

In the U.S., we use multiple-choice tests to test little bits of knowledge that are not directly related to the curriculum. (In fact, because curricula vary by state or even school district, companies that design standardized, multiple-choice tests pride themselves on divorcing their tests from curriculum.) Since the tests are supposed to be a surprise, going over questions from previous tests is almost like cheating. It's also a waste of time. Whatever little

bits of information the kids do learn have no context, so they'll be forgotten in a hurry. And a person looking at the test results will have no idea what they represent in terms of what the students know or can do.

Another disadvantage of not having a common curriculum is that we don't have any agreement on what teachers need to know. Colleges and universities can't train teachers on the basis of the curriculum they are going to teach, or assess them on how well they know it, because their students will end up teaching in many different school districts and many different states. What these students get instead are abstract courses that most teachers say were not even helpful in teaching them how to teach.

*An archive of Albert Shanker's weekly column in the New York Times is available at <http://nysut.org/shanker/>.*



or 9th grade (France), with students then streaming into separate educational tracks.

The existence of the national curriculum allowed for the creation of an entire education system geared to helping teachers teach the curriculum well. Teacher preparation and ongoing professional development were powerful because they were tightly focused on helping teachers understand the material they needed to teach and how to teach it. In most of these countries, *examinations* given toward the end of secondary schooling were based directly on the national curriculum or publicly distributed syllabi. Publishing companies planned their textbooks and supporting materials around the specific syllabi and curricula. Finally, the curriculum and syllabi themselves were typically easily available to the public; it was even for sale at regular bookstores. As a result, students, parents, and teachers all knew what kids should be learning; the possibility that expectations for poor and affluent students, especially in the lower grades, would

be quite different was greatly diminished (see box, p. 13).

## II. An American Educational System Based on Standards

Americans liked the coherence, alignment, and achievement results of these systems, but their centralization grated against the American tradition of local control of schools. The search was on to find a uniquely American way to capture the benefits of an aligned education system, without losing local control. A great national discussion ensued. Among the strong public voices advocating an education system driven by clear, high, transparent academic standards was AFT's president, Albert Shanker, who wrote on the issue many times in his weekly *New York Times* column (see box above). An influential paper by Marshall Smith and Jennifer O'Day (which began circulating long before it was published in 1991) described a potential American version of such a steering system.



In 1989, the discussion moved to the top of the American agenda, when the National Governors Association (NGA) hosted the President's Education Summit with Governors. The groundbreaking meeting endorsed the idea of national educational goals and a process for pursuing them that didn't undermine local control. From there the discussion moved to the newly established, bipartisan National Education Goals Panel (whose Resource Group on Student Achievement was chaired by Lauren Resnick, this article's lead author), and then to the congressionally authorized National Council on Education Standards and Testing, which included elected leaders from both parties and private individuals from the worlds of education (including Lauren Resnick), business, and other fields. Following much debate, discussion, and compromise, a rough consensus emerged, as captured in the documents produced by these various groups, on the main elements of what has come to be

known as standards-based education. The basic tenets, which were further developed and honed in states and in federal legislation, included the following:

1. Use a public process—involving educators, parents, community members, and potential employers—to establish common and transparent expectations, known formally as standards, for what students should know and be able to do upon graduation and at certain key earlier grade levels.
2. Develop assessments geared to standards that students could prepare for and that could provide clear targets for teachers' instructional work with students.
3. To preserve local control, encourage districts and schools to enact instructional programs explicitly geared to the standards and to organize continuing professional development around those programs. Pre-service teacher training, too, was to be organized around the standards.

## Lack of Equity, Quality Push Standards Forward in '90s

In 1994, the U.S. Department of Education, under President Clinton, released a startling report that documented how much less learning was expected of children in poor schools than in other schools (OERI, 1994). Researchers examined the

math and English grades received by a sampling of students from poor and affluent schools and compared these grades with the students' actual math achievement using test scores from the 1988 National Education Longitudinal Study (NELS:88). They

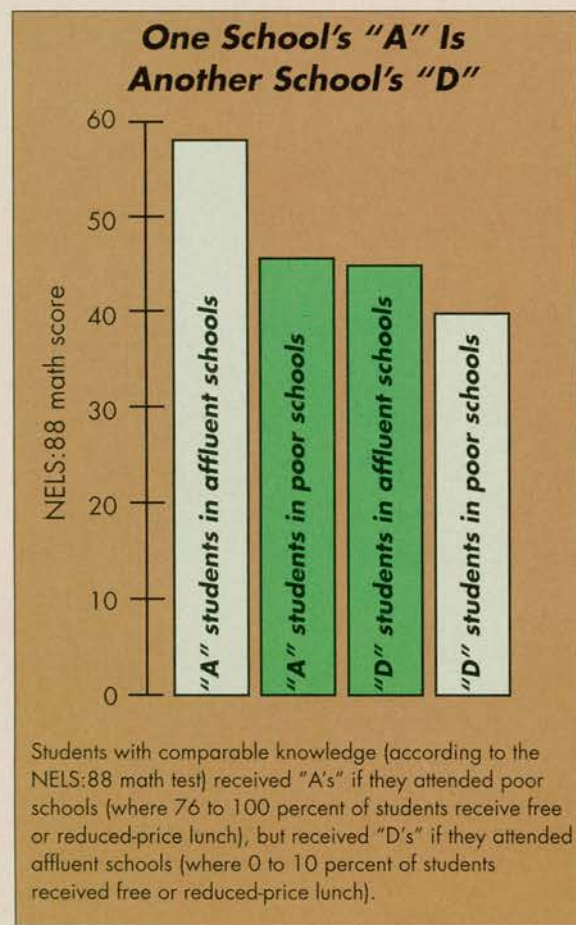
found that, on average, students with the same knowledge of math earned a "D" if they attended a low-poverty school—but earned an "A" if they attended a high-poverty school. (Results were similar for English.) In short, students in high-poverty schools were held to lower standards than were their middle-class counterparts.

Then in 1995 came TIMSS, the Third International Math and Science Study, which compared student achievement in 41 countries (Beaton et al., 1996). On the 8th-grade math assessment, 25 countries met the study's methodological requirements. Of these 25, U.S. achievement was surpassed by 14 countries, including all the Asian

and about half the European countries. News stories were quick to point out that the countries we "beat" were the vastly poorer Lithuania, Cyprus, Portugal, and Iran.

Concerns about the lack of equity and quality among America's schools weighed heavily on the minds of governors, especially in the poorer South. Standards-based reform received an additional boost from RAND researchers David Grissmer and Ann Flanagan's reports (1998, 2000) showing that the two states with an early commitment to standards and accountability—Texas and North Carolina—were posting the greatest gains on NAEP. Grissmer and Flanagan reviewed NAEP data from 1992-1996 and found that, when controlling for demographic factors, North Carolina and Texas had "greater combined student achievement gains in math and reading than any other states." According to the researchers, in addition to having in place such prerequisites as pre-K and smaller classes for low-income students, "the most plausible explanation [is] found in the policy environment ... the keys ... include[d]: creating an aligned system of standards, curriculum, and assessments; [and] holding schools accountable for improvement by all students."

—EDITORS





4. Create accountability systems that are based on whether students are meeting the publicly set and assessed standards.

The idea was that a standards-based system could combine the positive aspects of centralized curricula with the individuality and energy of the American local control system. The standards and assessments would be set by public entities such as states, but the details of curriculum, teaching, and professional development would be left to districts and schools. The accountability systems, rather than detailed regulations, would structure the priorities of schools and districts and press them to make the changes necessary to deliver effective teaching to all of their students.

It was an imaginative effort to harness the power of alignment without diminishing local control. It's also now clear that the task left to schools and districts—to create their own curriculum and instructional programs *and* figure out how to reinvent themselves to effectively deliver those programs *and* do it quickly—was enormous. The capacity of the schools to dramatically improve education was quickly outpaced by the much faster moving development of assessments and accountability systems. And this created the difficulties mentioned earlier: the inadequate support for teachers to meet ambitious new educational goals, the excessive focus on test preparation—in fact, in many places, the virtual hijacking of standards and education by narrow tests.

How did we get here—and how can we get back to the original intent of the standards-based system? To answer these questions, we'll look first at the development of standards; second, at the difficulties that have been confronted in bringing standards-based assessments to life; third, at the ways in which curriculum and professional development have (or haven't) been built around these standards; and fourth, where the rubber really hits the road, the accountability rules brought to us by the No Child Left Behind Act, signed into law in 2002.

## Standards

We begin with the academic standards. Who would write them? How detailed would they be? The years following the National Governors Associations' (NGA) 1989 Summit were a time of ferment as states, associations of states, the federal government, professional societies, non-profits, school districts, and individual schools all set about writing standards. For a time, it looked as though the lead role might go to national professional associations, such as the National Council of Teachers of Mathematics (NCTM), which in 1989 had written the first set of home-grown national standards. But there were also large states that developed their own standards or "curriculum frameworks"; one of the first in this early generation of standards documents was California, which launched a set of curriculum frameworks starting in 1987. And, there were efforts, such as that of the New Standards Project (which Lauren Resnick co-directed with Marc Tucker) to bring together consortia of states to prepare standards and related assessments (Viadero, 1994).

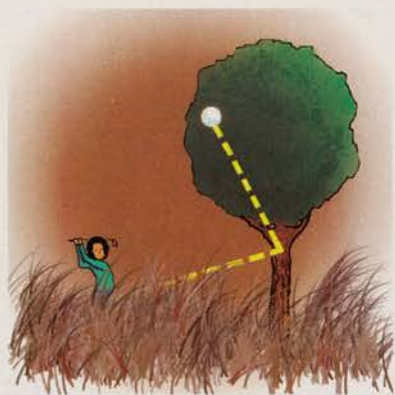
The first clear inkling that states would end up as the main makers and adopters of standards came in 1994. In that year, President Clinton signed the newly revised Elementary and Secondary Education Act (ESEA), renamed the Improving America's Schools Act, which required states to set statewide academic standards for its Title I students *that were the same as the standards that existed for other students*. This, of course, required any state that had not yet adopted standards to do so. The trend toward state standard-setting was locked in when No Child Left Behind (NCLB) became law in 2002. NCLB was yet another revision of ESEA, this one backed by President George W. Bush, with bipartisan congressional support.

Though the debate over who should set standards had subsided, there remained a question of just what standards should look like. How general? How specific? Should there be separate standards for every grade or should standards be specified just for broad "grade spans," such as 1-4 or 5-8?

<b>STRONG STANDARDS</b>	<b>vs.</b>	<b>WEAK STANDARDS</b>
<b>SCIENCE</b>		
Describe how groups of elements can be classified based on similar properties, including highly reactive metals, less reactive metals, highly reactive nonmetals, less reactive nonmetals, and some almost completely nonreactive gases. (Grade 8)		Describe the historical and cultural conditions at the time of an invention or discovery, and analyze the societal impacts of that invention. (Grades 5-8)
<b>SOCIAL STUDIES</b>		
Describe major rights, such as freedom of speech and freedom of religion, that people have under Indiana's Bill of Rights (Article I of the Constitution). (Grade 4)		Students will trace patterns of change and continuity in the history of their community, state, and nation and in the lives of people of various cultures from various periods. (Grade 4)
<p><i>Above are examples of "weak" and "strong" standards, as evaluated by AFT Educational Issues staff. The AFT's annual evaluation of state standards was published as Making Standards Matter, from 1995 to 2001. Since the late nineties, the AFT's reviews have been published annually by Education Week. For more information, including AFT's current state-by-state analysis of standards, see <a href="http://www.aft.org/topics/sbr/">www.aft.org/topics/sbr/</a>.</i></p> <p style="text-align: right;">—EDITORS</p>		



With the addition of accountability—and without a curriculum that defines broader educational goals—narrow tests may not serve simply as a floor, but can become the de facto curriculum. In short, the tests can hijack the rest.



States that chose to set periodic rather than grade-by-grade standards did so for what seemed to be a good reason—a widely held view among policymakers and others that we did not want people from outside the local school district to control every step of the curriculum. The same thinking led some states, whether their standards were grade-by-grade or periodic, to develop very general standards, instead of more detailed ones that approached the specificity of a curriculum (although there were also several states whose standards were so specific and lengthy that they were impossible to actually teach). But this view, that standards should be vague and/or periodic, ran into trouble, as we will see, as assessment and accountability developed.

While standards have gone through more than one round of revision in most states, they continue to vary widely in style and quality. Recent analyses of the overall quality of standards show a mixed picture and sometimes fail to agree on which states have good or bad standards (Stotsky and Finn, 2005; Klein et al., 2005; *Education Week*, 2004). (For examples, see box, p. 14.)

Clearly, the low-quality of some states' standards is a major barrier to realizing the potential benefits of standards-based education. Good standards are the foundation for the other elements of standards-based reform: a rich curriculum that builds important knowledge and skills in a logical sequence, professional development that focuses on teaching the curriculum, and assessments that measure whether students are reaching the standards.

### Assessments Aligned to Standards

Standard-setting was the crucial first step in building a standards-based education system. Next, in terms of attention and importance, was a new function for testing and assessment. Standards-based assessments were meant not just to judge performance by students and teachers (an accountability function, which we will discuss later), but also to serve as guideposts for teaching and learning. The idea was to create assessments that students could prepare for and that teachers could legitimately prepare students to do well on.

The idea of assessments designed to be taught to and studied for was new to most Americans (though in New York State, the Regents exams were of this type). But it was an idea familiar in most other developed countries that had for decades been using public examinations both as a basis for granting secondary school certificates and for university entrance (Resnick and Resnick, 1990). In examination-based education systems, it is normal and appropriate that curriculum and teaching are related to exams and aimed at helping students do well on them. In most European and Asian countries, for example, secondary school students take subject-matter examinations that are directly linked to a publicly specified curriculum. In some countries the exams are graded centrally by teams of teachers; in others, teachers grade the exams in their own schools and a sample of papers are graded centrally in order to “calibrate” local scores (so that grades coming from different schools, or even different cities, are comparable and students everywhere benefit from common expectations).

A crucial feature of these examinations is that students are rarely surprised by them. Both teachers and students know what to expect, indeed teachers draw on past exams as instructional guides. Not everyone likes all of the questions and study tasks, but teachers and students view the system as fair. What is more, external exams of this sort have the effect of turning students and teachers into a “team,” jointly working towards exam preparation. Similar teaching is seen in the U.S., when teachers prepare students for such externally developed exams as the Advanced Placement, International Baccalaureate, or some statewide exams, rather than their own end-of-course tests (Resnick and Resnick, 1992).

Examinations of this sort can take multiple forms. They can be “on demand” assessments in which students respond to set questions, including multiple choice questions, short constructed responses, extended essays, “performance assessments,” or extended pieces of student work produced over a longer period of time (“portfolio assessments”). An ordinary-looking test or an open-ended performance task becomes an examination when it is explicitly aligned to the curriculum or standards that students are meant to learn. Teaching toward well-constructed examinations is good professional practice.

Unfortunately, the tests that most states adopted to judge student progress toward state standards were not of this sort. Some states used the same tests (sometimes in adapted forms) that for years they had been purchasing from American testing companies, and these were not designed as exams. They were not systematically aligned to a specific curriculum or to standards that established what students should learn. In-



stead, they were designed to compare students with each other—spreading them out on a “bell curve.” The most common way of describing how much students knew, based on these tests, was to declare their “percentile” scores. Typically, being “at grade level” simply meant that you were at the 50th percentile—half of the norming sample scored higher than you, half lower. To make the tests work this way, test developers collected large pools of items that were thought to sample the average curriculum in use in American schools, tried them out on large populations of students, and then performed sophisticated statistical analyses to pick out the items that best “discriminated” among students—that is, spread them out on a normal curve. This was a far cry from constructing a standards-referenced or curriculum-referenced exam, in which one started with what one expected students to learn and developed test questions (or performance tasks) explicitly to match the standards or curriculum.

Many teachers have objected to being pressured to teach to norm-referenced tests, and indeed teaching to these tests is a bad idea. They were not designed to be taught to. Because they were meant to be used by many different school systems using many different curricula, they were not aligned systematically to anyone’s standards or teaching programs. In addition, because these tests depended on spreading students out on a curve, test items were retained or omitted in a test based on how they discriminated among students, not on how well they represented the standards to be taught. For all these reasons, it was impossible to tell from typical norm-referenced tests whether students were actually learning to expected standards.

Unfortunately, the problem of weak tests, not fully aligned to standards, is not limited to recycled versions of “off-the-shelf” tests. Even states that have constructed their own tests, based on their own standards, have largely relied on traditional test items and low-cost methods of scoring.

## Standards-Based Reform Brings New Attention to Key Elements Necessary for Improving Student Achievement

Standards-based reform and accountability have helped bring focus and attention to key elements necessary for improved student achievement, especially among poor and minority students in schools with the lowest levels of student achievement.

■ **Investment in early childhood education is up:** Poor children come to school already far behind their middle-class peers (NCES, 2001). Policymakers who want to raise standards and require students to pass tests for promotion or graduation have realized that they will have to invest more heavily in quality early childhood experiences. State expenditures on early childhood education have increased from \$267 million in 1988 to \$2.54 billion (in constant dollars) in 2002-2003 (Barnett, 2005; Barnett et al., 2004). Unfortunately, as states face financial problems, they often cut these very programs.

■ **Investment in early reading is up:** Likewise, there’s a growing understanding that students will not meet increased high school graduation standards unless they’ve received excellent early reading instruction. In the last two decades, enormous advances have been made in defining effective read-

ing instruction. First under President Clinton, with the Reading Education Act, and now under President Bush, with \$1 billion in 2005 for the Reading First Act, the federal government is bringing the new knowledge to the nation’s teachers. Many states (Texas, Maryland, California, Ohio, and Florida among them) have beefed-up their own investments in reading as well.

■ **Lawsuits requiring adequate funding are increasingly successful:** For the 15 years from 1973 until 1988, only seven of 22 such lawsuits were victorious—but since 1989, 19 out of 29 have been successful (Vock, 2004). According to Michael Rebell (2004), executive director of the Campaign for Fiscal Equity, “It is not a coincidence that the implementation of standards-based reforms and the accelerating plaintiff successes in the education adequacy litigations have occurred almost simultaneously since 1989.... [T]he new state standards for defining and assessing educational achievement have provided courts with judicially manageable criteria for implementing workable remedies in cases where the courts have invalidated state education finance systems.”

■ **Public confidence in schools is**

**rebounding:** In the eighties, confidence in public schools was low, according to the annual surveys published in *Phi Delta Kappan* magazine. In 1983, 31 percent of survey respondents gave their local public schools an “A” or “B” (White, 1983); in 1998, it was up to 62 percent (Rose and Gallup, 1998); and in 2002, it was up to 71 percent (Rose and Gallup, 2002). According to Public Agenda, which has tracked public views of education for over a decade, “surveys suggest that attitudes about local public schools have actually improved from 1998-2002—at least in the academic arena. Both professors and employers are less likely to say that local schools ask too little of students” (Johnson et al., 2003).

■ **Teacher frustration with lax academic standards has greatly decreased:** Polls of AFT teachers conducted by Peter D. Hart Research Associates in the early nineties showed significant dissatisfaction with low standards, including, for example, 46 percent of teachers saying that in 1994, they felt pressure to “pass students on to the next grade who really are not ready”; in the same poll, nearly one-third (30 percent) felt pres-



But in a standards-based education system, everything depends on how well assessments actually represent the full range of standards, in both topical content and cognitive demand—and thus on what kinds of teaching and learning behavior they evoke. Unfortunately, most state tests are not well aligned to state standards. In some extreme cases, alignment between state standards and tests is so weak that the standards from one state more closely match the tests used in another state (Porter, 2002). Most state tests do not do a good job of assessing the full range of standards and objectives that the states have laid out for their students. In fact, research has found that what “is included and excluded is systematic: the most challenging objectives are the ones that are under-sampled or omitted entirely... [and those] that call for high-level reasoning are often omitted in favor of much simpler cognitive processes” (Olson, 2003). As a result, although most state standards explicitly call for conceptual understanding and problem-solving, their tests often fail to

assess these standards. When teachers match their teaching to what they expect to appear on state tests of this sort, students are likely to experience far more facts and routines than conceptual understanding and problem-solving in their curriculum.

One could argue that with these tests we could at least measure whether students were acquiring the basics and that, for some students, a concerted effort to assure that they acquire the basics represents an improvement. But, as we will see, with the addition of accountability—and without a curriculum that defines broader educational goals—narrow tests may not serve simply as a floor, but can become the de facto curriculum. In short, the tests can hijack the rest.

### Curriculum and Professional Development Aligned to Standards

A strict test-based accountability system invites this kind of test-matching behavior. In theory, it is the standards that

sure to “give higher grades than students’ work deserves” and to “reduce the difficulty and amount of work you assign.” But that’s changing. Between 1994 and 2002, Hart’s polls found that the percentage of AFT teachers who believed that academic standards were too low dropped dramatically—from 51 percent to just 23 percent. (At the same time, there is increasing dissatisfaction with aspects of reform, including the number of tests given and the time devoted to test preparation.)

■ **More struggling students are receiving special interventions:** Where states and cities have established clear proficiency standards for promotion to the next grade and for graduation, there has often been a flow of resources and attention to providing interventions for struggling students. In Chicago, for example, when social promotion was ended, the school district created after-school and summer programs (with small classes of about 16 and a specially developed curriculum) for the thousands of students who were in danger of being retained (Roderick et al., 2003). In Charlotte-Mecklenburg, low-achieving students may be assigned to smaller classes, special tutoring, double doses of reading or math, or afterschool or summer programs (Snipes et al., 2002).

Similar efforts are in place in states such as Virginia and Massachusetts, where students must pass exams to

graduate from high school. In Massachusetts, for example, in 2001, just 68 percent of all students—and only 37 percent of black students—passed (on their first try, as sophomores) newly required high school exit tests in math and English. The state, school districts, and non-profit groups have worked to provide failing students with intensive assistance to master the content included in these tests. By the time these students were seniors (in 2003), 95 percent, including 88 percent of blacks, passed both tests (Achieve, 2004). But the state resources for these special assistance programs have recently declined, leading advocacy groups to argue for more resources for pre-K, afterschool, and other programs. Further, according to AFT’s *Making Standards Matter* report (2001), half of states either didn’t require or didn’t fund interventions for failing students.

The Center on Education Policy (2005) found that among school districts with schools that failed to make AYP, 99 percent (according to district self-reports) were providing “extra or more intensive instruction to low-achieving students”; 84 percent were providing “before- or after-school, weekend, or summer programs”; and 48 percent were hiring “additional teachers to reduce class size.”

Across the country, the students getting these extra services are the very

students who, absent these accountability requirements, were neglected in the past. However, according to CEP, just 20 percent of districts with the neediest students say they have adequate money to assist schools identified for improvement under AYP.

■ **Knowledge about improving achievement in low-performing schools is growing:** The standards movement’s clear focus on achievement is pushing new investments in researching the effectiveness of specific curricula and is also spawning a great deal of new knowledge about how to help specific schools improve themselves. Examples include the press for schools to use research-based methods (exemplified by the U.S. Department of Education’s recently established Institute of Education Sciences), some states’ formation of school assistance teams (like the ones that North Carolina created in 1997-98), the research community’s drive to develop replicable models for comprehensive school reform (see, for example, the *Catalog of School Reform Models* at [www.nwrel.org/scpd/catalog/index.shtml](http://www.nwrel.org/scpd/catalog/index.shtml)), and some think tanks and associations’ attempts to distribute reliable information on school improvement research (such as the RAND Corporation’s Promising Practices Network, online at [www.promisingpractices.net](http://www.promisingpractices.net)).

—EDITORS



teachers should be aiming for, but it is the far narrower tests that carry the consequences. Many principals distribute practice material designed to prepare students for the tests; and commercial test prep materials, billed as diagnostic and useable as a basis for differentiating instruction, can be bought easily from various publishers.

That kind of test-driven teaching was not the goal of the standards movement. According to the vision put forward by the Goals Panel and the National Council on Education Standards and Testing, school districts would develop rich instructional programs with strong content and good pedagogy that would be explicitly aligned with state standards (not tests). The system would be a *coherent whole*, its practical functioning boosted by ongoing professional development. Tests would be part of that whole, but they would be grounded in the standards. And the specifics of how students would be taught the standards would be left to local decision-makers.

This element of local decision-making was the major factor that made an American standards-based system different from the national curricula used by other countries. But is it working? The goal, remember, was to produce the benefits of Europe's and Asia's nationalized curricula—a common, transparent curriculum for all kids; a basis for powerful, focused pre-service and professional development; quality textbooks and curriculum materials; and an assessment system that would enable teachers, parents, students, and the country to measure students' progress toward mastering the curriculum—without actually producing a national curriculum.

To realize these benefits, someone has to create a curriculum or standards specific enough to carry the load that is carried by other countries' national curricula. It could be the state, the district, or an independent group. But without a common curriculum to serve as the anchor, standards-based reform cannot produce the aligned system of professional development, textbook and curriculum materials that was promised.

The bad news is that, as Achieve noted in a 2002 report, most states have not provided teachers or others with clear curricular guidance. According to an earlier 2001 report by the American Federation of Teachers, *Making Standards Matter*, only nine states had in place even half of what was necessary to provide teachers adequate curriculum guidance.

At the district level, the news has not, until very recently, been inspiring either. Our Institute for Learning works with some of the urban school districts that are trying the hardest to raise their students' achievement. Our work often begins with a "stock-taking" that includes examinations of test data coupled with classroom visits and discussions with teachers aimed at understanding the ongoing teaching program. We ask, "What is your curriculum?" Until recently, in most of these districts, both teachers and administrators described their curriculum by naming a textbook. Further discussion revealed that rather than defining a coherent program or assuring a common curriculum for all, the textbook was treated as a resource from which teachers could pick and choose materials for lessons, often adapting the material for their students. Teachers often did not know what their neighbors—

teaching the same grade, and the same course, and similar students—were doing with the adopted textbook, or even whether they were seriously using it. Consequently, students often experience a very fragmented program over the course of several years, a situation that is particularly negative when students (and even some teachers) change schools frequently. De facto, then, there often was no coherent curriculum, even within individual schools. Thus, the foundation of poorly aligned standards and tests is now overlaid with weak curricula—leaving teachers and the educational system with no common anchor except the tests.

We see hopeful signs that things are beginning to change, however. A number of districts, especially urban districts with mobile student populations, are beginning to recognize that a common curriculum across schools is a necessity. To boost student learning, some districts are also realizing that they need to greatly strengthen professional development, giving teachers the knowledge and skills they need to successfully teach challenging student populations that in the old days were expected to put in their seat time but not learn much. These districts are also realizing that effective professional development is based on a particular curriculum; it's not general and vague. In short, effective professional development requires the adoption of a curriculum; and the effective use of the curriculum requires ongoing, classroom-based professional development for teachers.

In response, many districts are going beyond merely adopting textbooks to implementing more fully "designed" curricula. They sometimes adopt programs designed outside the district (e.g., Open Court Reading or Everyday Math). Sometimes, they build district-wide instructional guidance systems that may use a textbook as a base, but add substantially more pedagogical guidance. These instructional guidance systems go well beyond the old "scope and sequence" charts that mainly suggested a flow of content. The new guidance systems can specify sequences of topics, suggest specific instructional practices both from a textbook and "supplementals" (or classroom libraries), the amount of time each topic should take, curriculum-embedded assessment tasks, student work samples, and sometimes, model lessons for use in professional development. Although these instructional curricula are sometimes tightly defined, all of those where we have seen achievement increases specify a mix of conceptual and skill emphasis. None call for teachers to read a "script" to students or to expect preprogrammed answers from them. All depend on providing intensive professional learning opportunities for teachers. These positive results of linking professional development to a specific teaching program are what we might expect given the growing body of research demonstrating that academic achievement increases when professional development focuses on the specific content teachers are expected to teach (Cobb et al., 1991; McCutchen et al., 2002). In one study, for example, David Cohen and Heather Hill (2001) found that most teachers who reported improved instructional practices had attended substantial training programs focused specifically on the curriculum materials that they used in their classroom. Those teachers' schools also posted



---

To realize these benefits, someone has to create a curriculum or standards specific enough to carry the load that is carried by other countries' national curricula. It could be the state, the district, or an independent group.



---

higher scores on a state mathematics assessment. By contrast, other professional development programs showed no such effects.

What we see, then, is the beginning of an effort in a growing handful of districts to make standards-based reform realize its full vision, not just instruction narrowed to tests. And, where it is happening, student achievement seems to be responding. But the magnitude of the effort being exerted in these districts cannot be minimized. The work we've described is typically being undertaken in large urban districts with strong district leadership and community support, where the infrastructure and economies of scale exist to support the large-scale implementation of curriculum and related professional development. But even in these districts, assembling the resources and know-how has been a challenge. What will be necessary to help other less able districts to move in this direction? What about students in smaller districts that can't afford such an investment? States—probably with federal support of different kinds—are going to have to figure out how to bring curriculum guidance and professional development to a much larger population.

### **Accountability**

And so we arrive at a discussion of accountability. We've seen that the standards that exist around the country are of mixed quality, with many quite weak and vague. Layered on top of these weak standards are tests that are typically not well-aligned; and in almost all cases, the tests measure students' progress on basic knowledge and skills, but rarely on the higher-level cognitive abilities that are included in the state's standards. In many places, there is no detailed curricu-

lar guidance that would allow teachers across a district to teach a common curriculum that went beyond what was tested; and without this curriculum, obviously there is not the related training that would support teachers in teaching it.

If you layer high-stakes accountability atop all of this, the formula is complete for allowing a narrow test, focused on the lower end of the curriculum, to hijack broader educational goals. And indeed, in many places, that was beginning to happen even before No Child Left Behind (NCLB). But with the adoption of NCLB, the threat became nationwide. Whatever pressure already existed to teach to the test increased, both because the consequences imposed by NCLB were more dramatic and—due to NCLB's formula for defining whether schools had made “adequate yearly progress”(AYP)—because they affected more schools.

Any accountability system layered on such a weak foundation would cause problems. But NCLB has unique features that cause additional, unique problems. Among them: Its formula for judging whether schools have made AYP does not take into account where a school started or how much progress it has made, which means that schools that have made great progress (but not enough to make AYP) will nonetheless be identified as “in need of improvement” (see “The AYP Blues,” p. 35). The particular consequences that it prescribes and the order in which they are prescribed can mean that wrongly identified schools will be subjected to consequences that can impede their further progress and thus hurt their students.

Further, the requirement that everyone must be proficient by 2014, while meant to encourage states to set high expectations for all types of students, is in reality encouraging states to set lower standards for everyone: The lower the standard, the easier it is for schools to meet the targets and avoid sanctions. Leaving the standards up to states was, of course, among the political compromises that made NCLB possible. It is a “states' rights” and “local control” solution embedded in a national law. But it creates an incentive to lower, rather than raise, expectations. For example, Pennsylvania deliberately lowered its proficiency standards after too many schools failed to clear the AYP bar. Some commentators believe the current law is creating a “race to the bottom,” undoing years of gradual rises in expectations and achievement (Ryan, 2004).

But even layered on a weak foundation, accountability, as it has played out, whether under NCLB or under certain state and local systems, has a silver lining that should not be dismissed lightly. It has brought substantial attention to teaching core, basic skills to the lowest-performing students and to a variety of programs that are increasingly aimed at improving the lowest performing schools in a district (see box, p. 16). And, in the case of NCLB (and state and local systems that disaggregate test data according to minority and poverty status), it has brought a special spotlight, and needed instructional focus, to helping poor, minority, ELL, and special education students improve their performance on the narrow (but important) body of skills and knowledge defined by state tests.

*(Continued on page 44)*



# Building New Knowledge

## *School Improvement Requires New Knowledge, Not Just Good Will*

By Richard F. Elmore

*"The problem with most incentive structures is not getting people to do the right thing. It's getting people to figure out what the right thing is to do."*

—Thomas Schelling,  
Distinguished Professor,  
University of Maryland School of Public Affairs

In the next two or three years, a very large number of schools, most of them urban, with largely poor, minority student populations, will be classified as failing<sup>1</sup> under the accountability provisions of No Child Left Behind (NCLB). This classification will trigger a series of increasingly harsh sanctions, ending with the shutdown of the school and/or its transformation into a charter, the handing off of its management to a private company, its takeover by the state, or another comparable, federally approved governance change. The law provides little assistance to these schools in their run-up to being shutdown. And, as more schools get identified (and they will), there will be even less assistance available to any given school. This policy appears to be driven primarily by an extraordinary belief in the power of incentives. The logic seems to be: If schools are threatened with closure and other sanctions, they will figure out how to improve themselves.

*Richard F. Elmore is the Gregory Anrig Professor of Educational Leadership at the Harvard Graduate School of Education and a co-director of the Consortium for Policy Research in Education. He has written numerous articles, reports, and books, including materials for the Albert Shanker Institute. This article was adapted with permission from Richard F. Elmore, "Doing the Right Thing, Knowing the Right Thing to Do," School Reform from the Inside Out: Policy, Practice, and Performance, Cambridge: Harvard Education Press, 2004. Copyright © 2004 by the President and Fellows of Harvard College. All rights reserved. For more information, visit <http://gseweb.harvard.edu/~hepg>.*

My theory is that these and similar policies toward failing schools are based on either faulty knowledge about school failure or no knowledge at all. I am a great believer in incentives; I've taught students about their power in numerous courses over the years. So, I'm not going to argue that they don't matter. They do. Nor would I argue that schools can't improve, to a degree, by just getting people more focused, encouraging them to make better use of their existing capacities, and making them work harder. But I also know that Thomas Schelling, one of the most astute economic theorists on the subject of incentives, was right. He told me, when I was a graduate student, "the problem with most incentive structures is not getting people to do the right thing. It's getting people to figure out what the right thing is to do." This is the problem I would like to focus on.

In this article, I want to argue that while incentives are important, they won't be nearly enough to bring about the dramatic improvements in student achievement that we all hope for. The performance targets set by current accountability policies are, for many failing schools, completely unattainable using their existing capacities. The knowledge simply doesn't exist in these schools to make the huge leaps in achievement that the law requires. Further, I want to argue that if we don't provide school staffs with what is necessary to make these leaps, that is, the knowledge and tools they need to raise student achievement, we will not only assure that schools don't improve substantially, we will increasingly sow cynicism and resistance toward the law. Asking people to do the impossible without helping them to master the skills necessary to do it is a formula for political resistance and ultimate failure. (Remember the old Russian saying, coined during the decline of the Soviet system, "We pretend to work. They pretend to pay us.")

I want to start with descriptions of two schools that are failing according to their state accountability systems and are







very likely to end up on NCLB's "needs improvement" list. I think these portraits help illustrate the huge lift that is required to dramatically improve a school over a sustained period and the extent to which even motivated, intelligent staff lack the knowledge to bring this improvement about. From there, I will lay out a general framework for how I think schools do in fact build their capacity to improve. Lastly, I will suggest how policy could enable many more schools to build this capacity and, with it, substantially raise student achievement.

## Portraits of Two "Failing" Schools

### Thornton Elementary School

I am observing a second-grade classroom at Thornton Elementary.<sup>2</sup> The teacher is working with a third of the students in one corner of the classroom doing guided reading, a form of literacy instruction in which teacher and students jointly read aloud and discuss a book with an explicit focus on the author's meaning, as well as the readers' responses to the text. Guided reading is new to the teacher, a veteran of 20 years. She is concentrating very hard. The students are also working very hard and seem to be successfully reading and responding to the book. Each student in the class will rotate through guided reading in the course of the literacy block—the 90-minute period every morning devoted to reading and writing at Thornton—in one of three groups. While the teacher is focusing on the eight students in the guided reading group, the remaining two-thirds of students in the classroom—about 16—are doing a variety of things. Two reading specialists are working individually with two students, obviously struggling readers, on specific problems of phonics and word identification. A classroom aide is supervising a group of students who seem to be filling out worksheets. Some students are reading on their own and writing in journals. There are books in considerable quantity available to students. Student writing is prominently displayed on the walls. In general, the classroom appears to be orderly, quiet, and efficiently run. Behavior problems are few. Students seem compliant and relatively happy. Above all, the adults seem to be very focused, working hard, and highly motivated.

The principal and superintendent have worked out a professional development strategy for the school that focuses time during the school day and during designated professional development days on priority instructional areas. The teachers uniformly say that this is the best teaching they have done. The teaching force at Thornton is a veteran group; the least experienced teacher has been there 12 years.

Over the course of the morning I visit several classrooms at Thornton. Each looks roughly the same in structure and texture. To the casual observer, it would be difficult to see why Thornton is a failing school. Teachers are working hard. Students are highly engaged. There are extra adults to work with failing students. The classrooms and hallways are orderly and clean. Thornton certainly looks nothing like the stereotype that laypeople might carry in their heads about failing schools—chaotic, disorderly classrooms, teachers obviously out of their depth with both content and student

---

Thornton and Clemente are both improving schools by any reasonable definition, but they are both failing schools under the terms of the current accountability system.



---

discipline, low-level student work, etc. In fact, most observers would probably say that, overall, Thornton represents a strong and positive environment for students.

I have been invited to the school by the superintendent and the principal because, after some initial modest success on the state reading and writing test, Thornton's test scores have gone flat. The student population at Thornton is more than 80 percent poor, with equal numbers of African-American and Latino students. The school is in an economically depressed city, and the patterns of student performance at Thornton reflect similar patterns in other elementary schools in the district. The state reading and writing test is a challenging test for even the highest-performing schools in the state. For Thornton, it is daunting. The superintendent and principal report that teachers in the school are heavily demoralized by their designation as a failing school. They feel that they have given the school's new literacy program their best shot. They feel they have dramatically changed their practice. The changes they have made are clearly visible in all the classrooms in the school, they feel, but they are still not making progress against the standards of performance they are expected to meet. After implementing the new literacy program (which included in-class professional development for guided reading), they moved some students out of the lowest level of achievement, and they even increased the percentage of proficient students. But further improvement



has eluded them and, perhaps tellingly, they never saw any increase in the percentage of advanced students.

### **Clemente Middle School**

Shift now to Clemente Middle School, a school of about 1,000 students, grades six through nine, in a large northeastern city. Essentially all students at Clemente meet the income requirements for free or reduced-price lunch—the prevailing measure of poverty. They are predominantly Afro-Caribbean, Spanish-speaking immigrants, with a significant number of African-American students. A large proportion of the students are from families that might be classified as the working poor—they perform the basic services of the economy with very low pay. It is mid-morning, and I am observing a seventh-grade language arts class, taught by a novice teacher—a Teach for America corps member, one of several in this school. The teacher is a new graduate of a prestigious northeastern liberal arts college. She is young, energetic, highly engaged in her work. She is African American and obviously has a strong rapport with her students—about 15 of whom are in her class today. One student sits off by herself in a corner, focusing on something on her desk—probably some form of “time-out” discipline problem. The rest of the students sit comfortably at moveable desks, focused on the teacher. The lesson has to do with topic sentences and lead paragraphs, a key element of the state’s middle-grades writing test, which these students will be taking next year. The teacher is carrying on a lively discussion of a topic that students are asked to use as the basis for their writing. As students volunteer ideas and write them down in their notebooks, the teacher actively engages them in a discussion of what they will write and how they will write.

This classroom is one of a number I have observed this morning and the patterns are similar: active teachers, highly engaged students, instruction targeted at skills that are, at the same time, useful on their face and included in the state reading and writing test. I do not see a discernible difference between the novice and experienced teachers on these dimensions. The principal takes me to visit a couple of classrooms where he knows he has problems with the teachers. These classrooms are noticeably less engaging places for students, the teachers are clearly struggling with the fundamentals of teaching; they also seem aware that they are not doing great work.

Clemente has four assistant principals, each of whom has instructional and professional development responsibility for a grade level in the school. The assistant principals are clearly present in classrooms. The principal and assistant principals have a strategy for professional development in key subjects with teachers. While time is limited, teachers participate and say the work is valuable to them in the classroom.

Clemente Middle School is a vibrant and exciting place visually. It is a relatively new building, with a large atrium as a central feature. It has a privately funded arts program in which students produce stunning examples of visual arts and writers from the neighboring city visit, while teachers and students conduct author studies of their work. The building exudes energy. Student work is visible everywhere, especially in the atrium.

Again, I am in the school because performance, after a brief gain, has gone flat—well below the target level required to keep the school from being classified as failing. The superintendent and the principal want me to see, hear, and feel what the school is like, not just examine the test scores. The superintendent thinks the principal of the school is one of the best in the district and is worried about losing him to a neighboring district with much higher-performing schools. The principal says, in passing, that he has had to learn to ignore much of the feedback he gets from the state in order to focus on the things that need to be done to improve the school. Teacher turnover in the school and district is about 15 percent per year. About 40 percent of the teachers in the district have four or fewer years experience. Virtually none of the Teach for America members stay after their two-year term is over. The district invests heavily in professional development in literacy and math, but the superintendent says that once the new teachers have received the basic staff development program, they are attractive recruits for neighboring suburban districts that offer them significantly higher salaries.

Again, to the lay observer, Clemente would not be thought of as a failing school. Although it has at least its share of marginal teachers, what you see as you walk the halls and visit classrooms are powerful examples of students doing interesting and creative work, teachers working hard to engage students in learning that is clearly connected to what the state tests measure, and students largely responding in the ways teachers want them to. Take away the discouraging test scores, and you have a school that most lay observers would say is a decent place for kids to learn.

I want to stress that based on my visits to failing schools in several localities, Thornton and Clemente are *not* atypical of many of the schools that are, and that very shortly will be, classified as failing under NCLB. These schools have been the object of intensive efforts to make them work better. People in these schools—teachers, administrators, students—are aware that they are in organizations labeled as failing, and, with certain exceptions, they are not happy or complacent about it. Liberal critiques to the contrary, failing schools are usually not resource-poor environments. They are heavily staffed, they have large numbers of specialists who work directly with students, and they have considerable access to outside guidance and expertise in most settings. They also frequently have access to community resources that bring considerable assets to the schools. Failing schools do not have uniformly weak leaders. Some do. Some don’t. The point is that “strong” leaders—as in the case of Thornton and Clemente—are often just as baffled about what to do about their situations as “weak” leaders, though strong, competent leaders may have more motivation and ability to find out what to do.

To be sure, I have also been in failing schools over the past several years that more closely resemble the common stereotype of such schools: schools that show little or no evidence of consistent expectations around the quality of instruction or student performance; schools in which the adults assign responsibility for low student performance to



families and communities rather than to themselves; schools in which the resources available to support student learning are managed in a chaotic and scattered way, if at all; schools in which teachers and students cannot answer the most basic questions about the purpose and direction of their work.

In general, what my colleagues and I have found in our research on accountability is that genuinely failing schools fundamentally lack what we have come to call “internal accountability.”<sup>3</sup> That is, they lack agreement and coherence around expectations for students in learning and they lack the means to influence instructional practice in classrooms in ways that result in student learning. In our research, high internal accountability leads directly to observable gains in student learning. Some failing schools lack internal accountability on anything but the most basic expectations—order in the hallways, for example. But, as the Thornton and Clemente examples illustrate, some failing schools are actually engaged in developing internal accountability and have had some success in generating increased student learning, but are still at risk of failure, and under the terms of No Child Left Behind, are likely to lose their franchise before they have an opportunity to meet the performance requirements of the law.

The conventional view that drives current policies regarding failing schools is that schools fail because they lack the proper incentives to succeed. These beliefs, I think, are embedded in accountability policies that focus on external rewards and sanctions as motivators for teachers, administrators, and students. These policies also focus on changes in governance and incentives (e.g. charters, school choice) on the theory that the “right” external incentive structure will “drive” schools and school systems to recruit and hire the “right” kind of people who will, in turn, lead schools toward the “right” kind of goals.

As I noted earlier, incentives matter. But one of the main insights I take away from my recent visits to failing schools is how clearly most of these schools have gotten the message that they are failing. The problem is that the message doesn't tell them what to do, other than to “get better.” This is the problem that is exemplified by Thornton and Clemente.

### **Thornton and Clemente: What to Do?**

As we debriefed our observations of classrooms at Thornton with the superintendent, the principal, and the lead teachers, a number of patterns became clear. First, while teachers were working hard to apply their new knowledge on literacy instruction, no one was paying attention to the overall instructional quality and intensity of what was happening to students while the teachers were doing guided reading. All one needed to do was to walk around the classroom and observe what students who weren't involved in guided reading were actually doing. Not much was going on with the other students when they were not in the group, with the exception of the two students who were working one-on-one with the reading specialists. But second, these two students presented another problem. The work they were doing with the reading specialists, while it was quite skillfully designed and done, was not explicitly connected to the work that these students were expected to do when they were not in remedi-

ation. That is, the intervention was successful, but it wasn't successful as a cumulative goal that would bring these students into the mainstream of the class. And third, when we asked students to describe to us, in real time, what they were doing when they weren't in a guided reading group, what its purpose was, and how they would know whether they had been successful at doing it, most of the students were unable to answer.

So it's not surprising that Thornton had some initial gains with its literacy program and then its performance went flat. What happened was that students were exposed to a potentially powerful reading intervention—the introduction of guided reading—that substantially increased the amount of time and the intensity of instruction for them relative to what they had been doing. This new activity got the teachers and students focused on coherent work around reading in a way that could be observed and improved. Clearly, the next increment in performance will come from increasing the level of intensity, cognitive demand, and coherence for all students, whether they're in guided reading or not. This will require the teacher to pay much more attention to the orchestration of activities in the classroom and to have much more clarity and agreement with students and support staff around the purpose of the work.

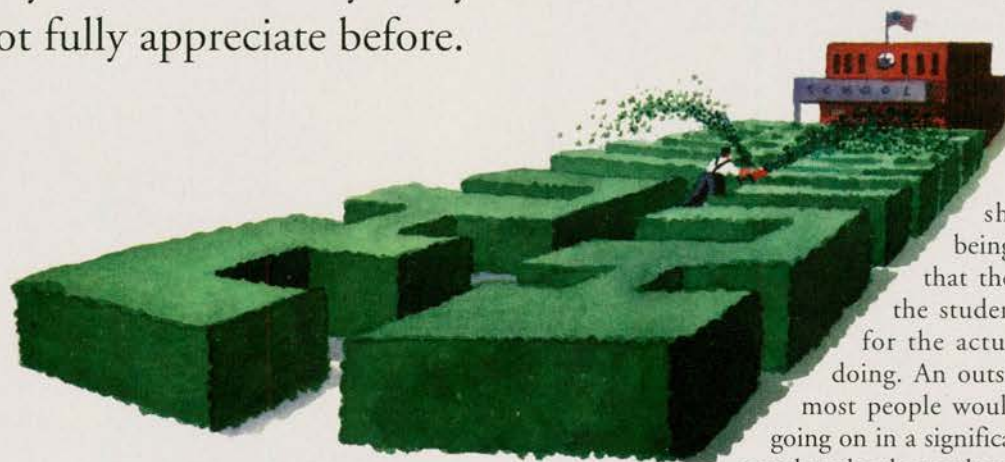
Several things are happening here. First, it often takes another set of eyes to see what principals, students, teachers, and support staff don't see, because they are working on solving the current problem, not on identifying the next one. At Thornton, the teachers were working so hard at mastering guided reading, they didn't have time to focus on what else was happening in the classroom. Second, teachers and students get more powerful in their practice, often against their own expectations, when they are brought to acknowledge a barrier and then put in the way of knowledge about how to get over it. Notice, it is important to understand that teachers and students don't get better by applying knowledge and skill they already have—they are stuck because their existing knowledge isn't enough. They get better by having access to new knowledge and discovering that they can use it in ways they did not fully appreciate before. Third, it is increasing the level of intensity, cognitive demand, and coherence around instructional practice that produces gains in student performance, and that process requires that everyone, including students, teachers, and support staff, develop increasing agreement about what the work is. This is what we have called internal accountability. If you walk into a classroom and sit down next to a student, ask him what he is doing and why, and you don't get a clear answer, it is highly unlikely that any powerful learning is taking place.

**A**t Clemente, as we debriefed, a different set of problems emerged. This will sound odd, but bear with me. The teachers at Clemente were working *too* hard. Novice teachers and veteran teachers of students in the middle and upper grades often equate “good” teaching with teaching that keeps students amused, interested, and seemingly engaged—which usually means eyes forward, paying attention, not causing any discipline problems, and respond-



---

It is important to understand that teachers and students don't get better by applying knowledge and skill they already have—they are stuck because their existing knowledge isn't enough. They get better by having access to new knowledge, and discovering that they can use it in ways they did not fully appreciate before.



---

ing in a timely way to the teacher's questions.<sup>4</sup> So what had happened at Clemente was that the "good" teachers in the building—including some novices and some experienced veterans—had adopted a style of practice in which the teachers were doing virtually all the work in classrooms and the students were doing very little. The teachers felt they were giving it their best shot, and to a layperson's eye they were doing even more than that. The students were engaged and amused, and they certainly weren't complaining. But when you looked at the classroom as a setting for *student work*, it was clear that not much was happening. A straight transcript of classroom discourse, for example, would reveal that, in order to keep students' attention focused on the front of the room, teachers were asking predominantly factual questions—questions that could be answered literally by the student pulling the information straight out of the text on the desk in front of them. When teachers did ask questions that required higher levels of cognitive demand—interpretation, argument, analysis—the overall pace of previous questions meant that waiting even a short period of time for a student response seemed like ages, so the teacher quickly

moved on to the next question before the students could fully engage in the previous one. The actual written work that students were being asked to produce—remember, this is a class aimed at preparing students to pass the state writing exam, which includes open-ended writing prompts—was likewise short and truncated, apparently because the teachers had made the judgment that the students needed, again, a faster pace with more concrete tasks, in order to stay engaged.

What was happening at Clemente is what often happens in the early stages of instructional improvement—teachers are developing rudimentary norms of practice designed to signal their collective commitment to students' success. They are, in effect, developing internal accountability around student learning and performance, albeit at a very rudimentary level. In the absence of careful and thoughtful analysis of the kind of practice that would lead teachers and students to be successful on a demanding writing exam, teachers were doing what they thought they should do—working hard, being enthusiastic, demonstrating that they can hold the attention of the students—without much thought for the actual work that students were doing. An outside observer would see what most people would regard as "good teaching" going on in a significant number of classrooms and wonder why the results weren't more impressive.

More importantly, teachers were generally *doing what they knew how to do*, rather than doing what was necessary to produce the result they were trying to produce. In the absence of specific guidance that what they were doing wasn't going to get them where they wanted to be, they would, other things being equal, continue to do what they—and many others—regarded as "good teaching," without recognizing that it was precisely that kind of teaching that was producing the performance they were disappointed with. Doing what they regarded as the "right thing" was not enough. They would have to figure out what the right thing was to do and then figure out how to do it.

### **Building the Capacity for Improvement: What's Needed**

Thornton and Clemente are both improving schools by any reasonable definition, but they are both failing schools under the terms of the current accountability systems in which they operate. They will both almost certainly be classified as failing schools under NCLB. They are not failing because the people in them don't, for the most part, recognize their limitations or fundamentally believe in the principles on which the accountability system is based. In fact, people who work in both schools *accept* that they are not doing as well as they should. This is why they have asked for help. Thornton and Clemente are not improving as much as they would like, I think, because of a fundamental design



flaw in current accountability systems: the failure of policy-makers to bring capacity-building measures into alignment with performance measures in the design of accountability systems.

As noted up front, the performance targets set by current accountability policies are, for many failing schools, completely unattainable using their existing capacities. Most schools, even nominally high-performing schools, couldn't do this work using their existing capacities. In order to meet these performance targets, schools have to develop successively higher capacities. Each new set of capacities speaks to the next level of problem. Each level of increased performance carries its own new set of problems. Each new level of capacity requires a period of consolidation. Acknowledging the gap between capacity and performance in accountability systems isn't, I repeat, *isn't* an argument for abandoning performance targets altogether. It is, however, an argument for a more knowledgeable approach to setting performance targets.

Improvements in school performance, as my colleagues and I currently understand them, probably take a form something like this:

- Schools recognize and internalize problems of performance by paying attention to evidence on student performance.
- They choose a proximate performance target—increasing reading performance, for example—and focus their work on improving their individual and organizational capacity to meet this target.
- If they succeed in choosing the right target and developing the initial knowledge and skill in teachers and students around that target, they typically see a modest bounce in student performance. Often, these initial moves, in very low-capacity schools, consist of very low-level changes: devoting a set number of minutes per day to teaching reading; realigning the curriculum so that the content that is tested is actually taught before the test is given; identifying students whose performance could easily be improved, thereby making the whole school look better, etc. I have come to call this the “low-hanging-fruit” stage.
- These improvements are real, but modest, and the capacity for further improvement is not there. But the critical moment here is that the school has decided to make some collective commitment to a goal that has to do with performance. This is the first stage of developing internal accountability.
- If the organization reads its performance well, it is at this stage that schools often try to tackle a more ambitious kind of instructional improvement. This improvement is often focused on the adoption of specific curricula and instructional practice. This stage almost always requires that the school receive some kind of external help, support, and professional development. Why? Because, by definition, people in the school don't know what to do, or they would have done it already. New practices take time to acquire and implement with any consistency. They also require people to organize and manage themselves around increasingly clear collective goals—another increment in internal accountabil-

---

Acknowledging the gap between capacity and performance in accountability systems isn't, I repeat, *isn't* an argument for abandoning performance targets altogether. It is, however, an argument for a more knowledgeable approach to setting performance targets.



ity. But schools that go through this phase almost always see gains in student performance, in part because they are learning to work together more powerfully, and in part because they are actually teaching different content in different ways. Just as predictably, performance tends to go flat again almost immediately. This is where Thornton and Clemente were in the improvement cycle when I visited. As noted above, performance goes flat because the problems of improving student performance are more complex than the strategies adopted at this stage can cope with.

- If the organization diagnoses this problem—and schools usually require some kind of external help to do this—it then has an opportunity to examine the barriers to continued improvement. Typically, the kind of problems that schools work on at this stage are either problems of increasing the consistency and cognitive demand of instruction or figuring out why the instructional strategies they adopted earlier work for some students and not for others. Simply diagnosing these problems and working on their solutions creates new capacity for collective action, bringing the next stage of work and improvement.



■ The problems of improvement become more complex and demanding as performance increases; the challenges to existing instructional practices and existing organizational norms become more direct and difficult. Often, schools go through some kind of crisis at around this time, where teachers and principals argue that the work has become impossible to do under existing resource constraints and that expectations set by external accountability systems are simply impossible to meet. This is a very tricky stage, because it is hard to argue that the demands of the external accountability system are reasonable—they often are not. But what teachers and administrators are also saying is that they simply don't have the capacity to make the next set of improvements, and they usually don't. I have come to call this the "impossible work" stage. The conditions for future improvement are present, but the capacities to make that improvement are not. It is critical for schools to receive high levels of support at this stage—to get help in diagnosing the next set of problems, to get help from people with expertise about problems of student learning and instructional practice, to broaden and deepen common expectations around high-quality instructional practice, and usually to see schools in similar circumstances that have managed to move through this stage. It is at this stage that the credibility of accountability systems, as systems of political authority, is tested. Here's why: *My* authority to command or induce you to do something you are not currently doing depends, in large part, on *your* capacity to actually do it. You may be motivated to do it. You may agree with me that it should be done. Or you may be willing to do it just because I have a legitimate grant of authority to require you to do it. But if you *can't* do it because you do not have the capacity to do it, then my authority is diminished because I have induced or required you to do something you cannot do. I can flog you harder, I can penalize you, I can threaten you, but I cannot make you do something you do not know how to do.

■ Schools that make it through these crises typically emerge as much different organizations—stronger, more coherent, with responsibilities more widely distributed and with much higher morale around student learning and much higher cognitive demand in the classroom. But they often have difficulty demonstrating that these changes are consequential because going through a crisis saps the energy and commitment of people while it is going on, and often performance goes flat during these periods. So just as the school is feeling that it has a much better handle on student performance, its results often look less impressive than they should. This occurs because the school has built the capacity for higher-level instruction but hasn't yet seen its full effects. This is where more concentrated work on instructional practice—not less—is important, because it is important that teachers and administrators understand that not only have they changed the way instruction occurs in the schools, they have changed their own capacity to take responsibility for and manage their school's response to pressure for performance. Again, external support and assistance around targeted problems of student learning help to reinforce the idea that everyone has gotten here by developing new knowledge and skills.

■ The next stage of improvement—one that very few

schools achieve, even nominally "high-performing" schools—is where the school collectively takes over the management of its own improvement process, teachers and students internalize the values of managing and monitoring their own learning, administrators model their own learning for teachers and students, and individuals are empowered to ask for the help they need.

This is a highly stylized map of the improvement process, in part because we need much more research to get a deeper picture of what actually happens as failing schools improve, and in part because Leo Tolstoy was right—each unhappy family is unhappy in its own way. Each failing school faces, in addition to the general problems of improvement, a specific set of problems rooted in its own context.

C current accountability systems aren't built to do what they are supposed to do—to push and support schools in getting better. The systems exhort schools and localities to provide support and professional development for schools in need of help, but don't actually invest in the infrastructure required to make sure that that help gets to the right schools at the right time with the right technical expertise. They heavily underinvest in the development of the knowledge and skill required to rectify the problems that failing schools face. The systems are generally unresponsive to the systemic problems that prevent resources from getting to schools—resources necessary for high-quality work—and tend to view the problems of all low-performing schools as essentially the same. The most discouraging aspect of current systems is that they ignore and undervalue the struggles of people like those who work in schools such as Thornton and Clemente, creating the expectation that students would be better off in other settings, without understanding that moving students around is essentially moving the problems of capacity from one set of institutions to another, without remedying the underlying problems of how to raise the capacity of the institutions where the children are to begin with.

Schools don't suddenly "get better" and meet their performance targets. Improvement is a process, not an event. Schools build capacity by generating internal accountability—greater agreement and coherence on expectations for teachers and students—and then by working their way through problems of instructional practice at ever-increasing levels of complexity and demand. Right now, virtually no infrastructure exists to provide continuous support to failing schools.

Building capacity in failing schools is going to require a lot of feet on the ground—people who know something about school improvement and who know what they don't know. As my analysis suggests, I would look for these people in what I have called "improving" schools (not in nominally "successful" schools because a large number of "successful" schools are not improving schools<sup>5</sup>), where faculty and school leaders have worked through several stages of improvement, and in improving school districts, where district-level personnel have gained real knowledge about the kind

*(Continued on page 47)*



# Curriculum First

## *A Case History*

By Roger Shattuck

*The great truths in education turn out to be half-truths in search of their other half.*

On Town Meeting Day in March 2000, some 400 legal residents of Lincoln, Vt., elected me to a three-year term on the board of Mt. Abraham Union High School, located in neighboring Bristol. A few days later, I took my oath of office and settled into a schedule of bi-weekly meetings in the school library. Comprising grades 7 through 12, the school serves around 900 students from five rural towns for an annual budget topping \$9 million under a board of 13 elected members. Mt. Abe belongs to the Addison (County) Northeast Supervisory Union district. In March 2003, I was reelected to the Mt. Abe board and also elected to the district board coordinating six local schools.

During the first year on the high school board I felt overwhelmed by the sheer volume of important and sometimes ominous decisions that engulfed the board. We dealt with school security, contract negotiations with teachers, selecting a new principal, Internet filtering, special education mandates, and preparing the all-devouring, seemingly self-propelled annual document called the budget to submit to voters. I thought I'd never catch on, never learn the names of all the key people plus the acronyms used to designate occult entities in the wonderland of education.

After 40 years of college teaching, I had no particular agenda to promote on the board. Principally, I was curious to find out what was actually being taught in this rural high

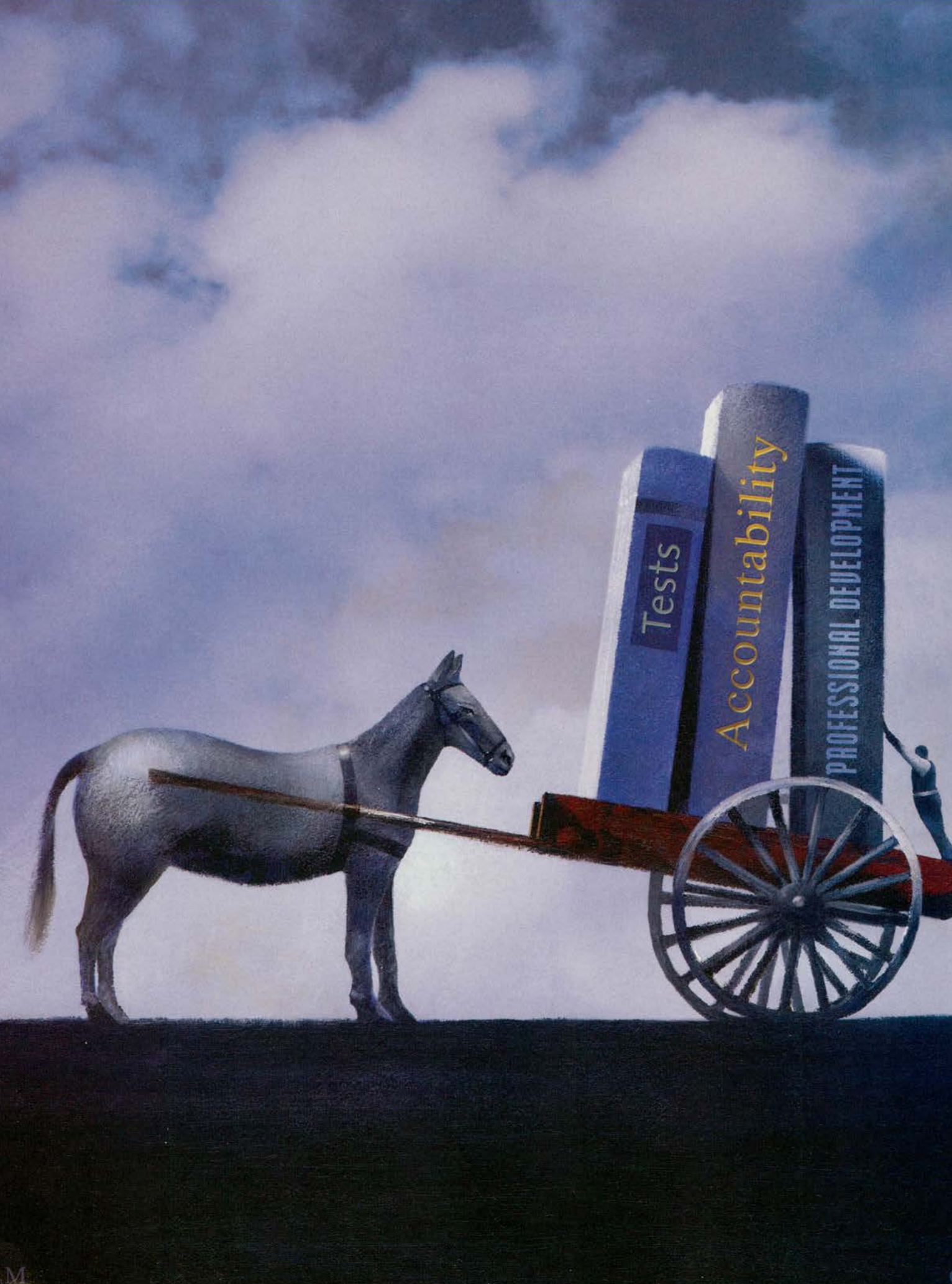
school, which has the largest payroll within 20 miles. I soon learned that the board spends little time discussing curriculum. I was told that the best way to inform myself would be to visit a few core courses. I chose English and History, or rather "Language Arts" and "Social Studies." (A return to the earlier names became the first item on my agenda.) Given a schedule by the department head, I visited about a dozen classes and was welcomed without fanfare or raised eyebrows. These visits gave me a vivid impression of overcrowding, of teachers without their own classrooms pushing overloaded carts like the homeless, of poorly and noisily ventilated classrooms, and of the constant demands imposed upon teachers for patience, firmness, and imagination. But amid multiple activities in the classrooms I found it impossible to discern a coherent sequence of content guiding the classes, not even in different sections of the same course. It would require months of class visits to gain an adequate sense of what was being taught in my school.

It turned out that there was another road to take. I had volunteered to be the school board representative on the teachers' curriculum committee preparing part of the self-study for our big 10-year NEASC accreditation visit. (NEASC stands for New England Association of Schools and Colleges.) This committee of eight teachers, chaired by the science department head, alerted me to four sets of documents dealing directly or indirectly with the curriculum. I made it my business to obtain and study all of these documents—as follows:

1. *Vermont's Framework of Standards and Learning Opportunities*, Vermont State Board of Education, 84 pages, 1996. (See the excerpt in the box on page 32.)
2. *Curriculum Guidelines*, Addison Northeast Supervisory Union. Six K-12 documents prepared and revised in rotation by the district in Mathematics, Language Arts, Science, Fine Arts, Social Studies, and Foreign Languages. They range in length from 33 pages (Mathematics) to over 200 pages (Language Arts), for a total of nearly 600 pages.

*Roger Shattuck is University Professor Emeritus at Boston University and author of many books, including *Forbidden Knowledge: From Prometheus to Pornography*. This article is based on an address he gave for a conference on the liberal arts sponsored by the Center for School Improvement at Boston University in May 2003. Earlier versions of this essay appeared in the *Journal of Education*, Vol. 184 (November 2, 2003) and in the *New York Review of Books*, Vol. 52, No. 6 (April 7, 2005), from which this article is excerpted with permission.*





Tests

Accountability

PROFESSIONAL DEVELOPMENT



3. *Course Selection Guide*, Mt. Abraham Union High School. Published yearly. Contains 60 pages of brief descriptions (3 to 15 lines) of all courses offered by the school.

4. Course syllabi filed by all teachers as required by law in the assistant principal's office. One to three pages following a recommended outline.

The first three of these documents form a stack over four inches high. The teachers on the accreditation curriculum committee brought none of the above documents with them to refer to. They kept borrowing my copy of the *Course Selection Guide*. All of the teachers appeared to acknowledge that document number 1, *Vermont's Framework of Standards*, contains the tables of the law. But they had not read it carefully. None of the teachers seemed familiar with or interested in document number 2, the district's own lengthy *Curriculum Guidelines*, prepared by committees of teachers meeting over a period of many months. The *Course Selection Guide* is little more than a useful list identifying all course offerings. The syllabi record what has been taught in a particular course or section, not a program of study approved by the school.

It is not easy to describe the first two official documents. The state *Framework of Standards* and the lengthy district *Curriculum Guidelines* (themselves based scrupulously on the state *Framework*) presumably lay out a course of study for all students. As they stand, however, these two documents do not and cannot serve this function. They mention no authors' names and no titles of books to be read. Only the science and mathematics documents specify topics for a particular grade. Elsewhere, entry after entry stipulates that students shall *examine, investigate, analyze, understand, and interpret* immense intellectual topics such as "fiction" and "nature and nurture." The verbs *teach, learn, and study* do not appear. Because they clump four grades together, these documents cannot, for example, provide an answer to the question: "In what grade are the following materials taught: the solar system, Athenian democracy, dangling modifiers, and the Founding Fathers?" Such items do not even appear.

The nearly impenetrable pages of the state of Vermont's *Framework of Standards* plus the Addison Northeast *Curriculum Guidelines* add up to an elaborate professional camouflage of the fact that at no level—state, district, or school—is there a coherent, sequenced, and specific curriculum. The teachers on the curriculum committee for accreditation had good reason to ignore the district *Curriculum Guidelines*—they propose no course of study, no coordinated sequence of subjects within the core fields. I'm not saying that our district curriculum is watered down or lopsided or old-fashioned or newfangled. I'm saying that those 600 pages contain no useful curriculum at all.

What then fills these pages in multiple copies that no one reads or consults? In large part, they contain bland hortatory statements about what students "should know and be able to do." It's almost a mantra. Yet, the two major curriculum documents refer to no specific content, to no simple lists of items such as osmosis or Martin Luther King, Jr. or, one hopes, Martin Luther.

And what also fills these pages, in the place of what to

teach, is lengthy instructions about how to teach these unspecified materials. Our district *Curriculum Guidelines* of recent years devote increasing space to "Best Practice in Teaching," identified as "an inquiry approach, which is based on constructivist principles." The documents to which one looks for the articulation of curriculum turn out to be presentations of a pedagogical doctrine, constructivism, which is much in dispute and has appropriated to itself the dubious slogan and sales pitch "Best Practice." Most of my fellow board members don't know what "constructivist" means and, if they read that far in the *Curriculum Guidelines*, they don't ask. (Constructivism refers to the half-truth that full understanding occurs only when students learn for themselves from hands-on experience without direct instruction or teacher intervention.)

I cannot draw general conclusions about American education from this description of Mt. Abraham Union High School and of the supervisory district it forms with five elementary schools. I have observed only this one case. But at conferences where I have presented some of these materials, other participants have not hesitated to respond, "That's a pretty good description of my district."

By going back to school as a board member, I have come to the conclusion that my school and its district have no ascertainable curriculum and no effective curriculum document. Various sources continue to provide topics to be taught—individual teachers, lesson plans, habit, informal consultation, tradition, inertia. Even without the guidance of a curriculum, education goes on. Teachers teach. Students learn. They may even study. Budgets are voted in. The caravan passes. But all is not well. Is there anything to be done?

## Dewey's Epiphany

During the two years it took me to discover the absence of an adequate curriculum at Mt. Abraham Union High School, I was also trying to reeducate myself about public education, elementary and secondary. I discovered all over again in books the intellectual excitement churned up by the history of education. That subject embraces the survival of democratic institutions, the conflicting claims of reason and religion, the nature of human cognitive development, the importance of personal leadership, and the constant distraction of intellectual fashion. (At the moment, we cringe or bask in the glare of several fashions: multiple intelligences, constructivism or discovery learning, personalized learning, and critical or higher-order thinking. They are all powerful half-truths.)

My reeducation in education led me to a curious discovery. At the turn of the 20th century in the U.S., John Dewey was conducting a famous yet now partially forgotten experiment in education at the University of Chicago, and he found himself in a position not too different from the educators at Mt. Abe.

After publishing four books in philosophy, psychology, and social thought and earning himself a sturdy national reputation, Dewey decided that, like any self-respecting science, the field of education needed a laboratory, an experimental setting, in which to develop and test its hypotheses. Dewey rapidly funded, founded, and staffed the University



Because they clump four grades together, these documents cannot, for example, provide an answer to the question: “In what grade are the following materials taught: the solar system, Athenian democracy, dangling modifiers, and the Founding Fathers?” Such items do not even appear.



Elementary School, soon to be known as the Laboratory School, with himself as principal, his wife on the staff, and their children attending. The school opened in 1895 with 16 students.

More seriously and consequentially than I have done by serving on a school board in Vermont, Dewey went back to school in his late thirties. He built and ran a laboratory, showplace, proving-ground school. What concerns us here is what Dewey himself learned by going back to school in this enterprising, private, yet essentially democratic experiment.

In 1897, Dewey displayed the reformist zeal that inspired his experimental school in a pamphlet called “My Pedagogic Creed.” Two quotes reveal his progressive approach:

The child’s own instincts and powers furnish the material and give the starting point for all education.

I believe that there is, therefore, no succession of studies in the ideal school curriculum.<sup>1</sup>

Two years later, Dewey published *The School and Society*, and one can see his progressivism beginning to incorporate some traditional ideas. In particular, one begins to hear two

new and unexpected verbs: to “direct” and to “control” the child’s activities.

After three more years at the Laboratory School, Dewey returned to the pamphlet format to share what he had learned in “The Child and the Curriculum.” Having observed the development of his Laboratory School for six years, Dewey now concludes that he wants it to be *simultaneously* child-centered and curriculum-centered. He is not proposing a compromise: He is promoting two complementary viewpoints. The logician in Dewey found the analogy of a continuum connecting apparent, not real, opposites. What he wrote deserves to be looked at carefully:

The child and the curriculum are simply two limits which define a single process. Just as two points define a straight line, so the present standpoint of the child and the facts and truths of studies define instruction. It is continuous reconstruction, moving from the child’s present experience out into that represented by the organized bodies of truth that we call studies.

Has Dewey now solved the problem of the child and the curriculum, either for 1902 or for 2005? No indeed. But something has happened. By going back to school, to his own school, Dewey allowed practice to guide theory to a sturdy synthesis. The Laboratory School under Dewey began to set, and then maintain, a year-by-year curriculum to guide the developing experience of the children. This major development in Dewey’s thought and practice points back to my original quandary and presents a balanced understanding of the role of “the organized bodies of truth that we call studies”—that is, a coherent curriculum.

Newly elected to the school board in Vermont, I had learned that in 600 pages of official documents, there was no attempt to lay out a curriculum—that the program of studies was essentially rudderless. I went on to explore Vermont statutes and regulations pertaining to curriculum. These perfunctory-sounding rules distribute power and responsibility among three parties: the state, districts, and individual schools.

The legislature has assigned to the State Board of Education the responsibility to set standards for student performance, and to the 60 supervisory boards in the state, the responsibility “to coordinate curriculum plans among the sending and receiving schools” in their districts. The *School Quality Standards* issued by the Vermont Department of Education stipulate: “Each school shall make continual and steady progress in the alignment of local curriculum consistent with the [State] Framework [of Standards] or comparable standards” and “each school shall evaluate and review the curriculum on a periodic basis.”

I interpret this overlapping legal language to mean that responsibility for setting curriculum lies with “each school,” subject to the “coordination” of the supervisory board and in accordance with state standards. Thus, the principle of local control remains unmentioned and presumably unchallenged. District boards can only “coordinate” the curricula of district schools; the state board can only establish “standards” for the schools’ curricula. Those state standards are deliberately kept vague, if not empty, for fear of infringing on local control.



# Excerpt from Vermont Reading Standards

PreK-4

5-8

9-12

## Reading Comprehension

1.3 Students read for meaning, demonstrating both initial understanding and personal response to what is read. This is evident when students:

- |   |   |   |
|---|---|---|
| <p>a. Comprehend grade-appropriate materials;</p> <p>b. Analyze and interpret features of a variety of types of text; and,</p> <p>c. Make connections among various parts of a text, among several texts, and between texts and other experiences in and out of school.</p> | <p><i>Evidence PreK-4 applies, plus—</i></p> <p>d. Make extensions/applications of a text;</p> <p>e. Identify the textual structure and/or the technical, artistic, and literary conventions of text; and,</p> <p>f. Explain the meaning of various forms of representation (e.g., narrative, graphical, cartographic, symbolic, mathematical).</p> | <p><i>Evidence PreK-8 applies, plus—</i></p> <p>g. Analyze, interpret, and evaluate texts produced for a wide range of purposes and audiences, including their cultural, political, and aesthetic contexts.</p> |
|---|---|---|

## Reading Range of Text

1.4 Students comprehend and respond to a range of media, images, and text (e.g., poetry, narrative, information, technical) for a variety of purposes (e.g., reading for pleasure as well as reading to develop understanding and expertise). This is evident when students:

- |   |  |  |
|---|--|--|
| <p>a. Read at least 25 books in a year, choosing high-quality materials from classic and modern literature and public discourse or their equivalent in magazines, newspapers, textbooks, media, and technical works;</p> <p>b. Read at least three different kinds (genres) of printed materials and at least five different writers;</p> <p>c. Read primary and secondary sources; and,</p> <p>d. Read at least four books (or book equivalents) about one issue or subject, or four books by a single writer, or four books in one genre, and demonstrate such reading through speaking, writing, or other appropriate means.</p> | <p><i>Evidence PreK-4 applies.</i></p> | <p><i>Evidence PreK-8 applies.</i></p> |
|---|--|--|

*An excerpt of reading "standards" from Vermont's Framework of Standards and Learning Opportunities, published by the Vermont State Board of Education. English courses at Vermont public schools are based on these standards.*

In practice, the result is an elaborate game of *"Après vous, Gaston."* Everyone gestures to the other parties to go first through the door of setting a genuine course of study. In my district, the *Curriculum Guidelines* in each content area are drawn up by a committee representing all five elementary schools, one high school, and the superintendent's office. Great pains are taken to make the *Guidelines* "standards-based"—that is, attentive to ("aligned with") the state *Framework of Standards*, as well as "coordinated" among different schools. In this elaborate ritual of deciding on the curriculum at three levels—local, supervisory district, and state—all specific content drains away. Scores of people in every one of the 60 Vermont districts spend weeks thinking up fuzzy professional language to compensate for the absence of a specific curriculum.

High over the general melee I have been describing hovers

the great raptor: I refer to the elusive yet commanding term "standards." No one can define them. No one can oppose them. No one can use them to write a curriculum. But is the mystery so great? In my understanding of our language, a standard (a required level of attainment in a defined activity) cannot exist in education without a curriculum to define the activity or field of study. We cannot "set the bar" higher or lower unless everyone involved knows the rules of the game and how to measure inches and feet.

Somehow, the laborious, confusing, 600-page "standards-based curriculum guidelines" drawn up for my Vermont district do not prevent a basic education from reaching a fair number of students in Vermont. Still, I'm convinced that we could do better and also save time and money.

All this I learned during four years sitting on two school boards.



## Excerpt from Core Knowledge English Curriculum—Grade 8

### Fiction, Nonfiction, and Drama

#### A. Short Stories

- “The Bet” (Anton Chekov)
- “Dr. Heidegger’s Experiment” (Nathaniel Hawthorne)
- “God Sees the Truth But Waits” (Leo Tolstoy)
- “An Honest Thief” (Fyodor Dostoyevsky)
- “The Open Boat” (Stephen Crane)

#### B. Novels

- Animal Farm* (George Orwell)
- The Good Earth* (Pearl S. Buck)

#### C. Elements of Fiction

- Review:
  - plot and setting
  - theme
  - point of view in narration: omniscient narrator, unreliable narrator, third person limited, first person
  - conflict: external and internal
  - suspense and climax
- Characterization:
  - as delineated through a character’s thoughts, words, and deeds; through the narrator’s description; and through what other characters say
  - flat and round; static and dynamic
  - motivation
  - protagonist and antagonist
- Tone and diction

#### D. Essays and Speeches

- “Ask not what your country can do for you” (John F. Kennedy’s Inaugural Address)
- “I have a dream”; “Letter from Birmingham Jail” (Martin Luther King, Jr.)
- “Death of a Pig” (E. B. White)
- “The Marginal World” (Rachel Carson)

#### E. Autobiography

- Selections from *I Know Why the Caged Bird Sings* (Maya Angelou)

#### F. Drama

- *Twelfth Night* (William Shakespeare)
- Elements of Drama
  - Review:
    - tragedy and comedy
    - aspects of conflict, suspense, and characterization
    - soliloquies and asides
  - Farce and satire
  - Aspects of performance and staging
    - actors and directors
    - sets, costumes, props, lighting, music
    - presence of an audience

#### G. Literary Terms

- Irony: verbal, situational, dramatic
- Flashbacks and foreshadowing
- Hyperbole, oxymoron, parody

*An excerpt from the Core Knowledge Sequence (1999), a curriculum that the author determines would be ideal for his school district. It is currently being used by nearly 500 K-8 schools and is published by the Core Knowledge Foundation.*

What Dewey learned by going back to school can be told more briefly. He knew more to begin with. He learned to acknowledge not one but two centers in school: both the child and the subject matter to be taught to the child. He found the fit between those two half-truths. Between 1896 and 1902 in Chicago, Dewey changed his mind and recognized the need for a coherent K-12 curriculum.

Furthermore, his conversion to a sequenced, specific curriculum throws light on a complaint often heard today about standardized tests: namely, that tests oblige teachers to teach to the test. But just reflect for a minute. The reason for teaching to the test is not the mandated existence of tests. It is the lamentable absence of a clear curriculum. *If there’s no coherent curriculum to teach to and to base tests on, then one has to teach to the test.* Here lies the great pedagogical short-circuit and breakdown, brought on by the empty promises and dummy documents called “standards.” Without a specific curriculum, there can be no standards.

### Finding a Curriculum

I don’t have to go back to the Greeks and Romans, or to the trivium and quadrivium, in order to make the simple point that today, the only way to assure sustained attention to a true liberal arts program in a school is to embed it in the

curriculum adopted by the school. A teacher here and there trying out Ovid or Dickens or a chapter of Tocqueville may ignite the intellectual curiosity of a few students and deserves every encouragement. But a curriculum that specifies a judicious selection of great books and perennial topics will allow that intellectual excitement to spread further and to attain the added rewards of “commonality.” Yet, I have visited class after class in which the choice of reading is left entirely to the student. Commonality of reading and study, not to be confused with lock-step, is neglected in favor of student choice and personalized learning.

Now, I am not so optimistic as to believe that my supervisory district will soon develop a genuine curriculum, and even a liberal arts program. I do not foresee that 65 teachers at Mt. Abraham Union High School, seconded by the district board and the Mt. Abe board and the teachers of the five feeder elementary schools along with their boards, will soon decide to draft a grade-by-grade, content-rich, specific, flexible, teacher-friendly, and teacher-proof curriculum—and then be able to adopt it.

Yet, I believe that the accompanying deliberations would stir up the school and parents in a healthy and fruitful way. I would love to hear members of our community discussing the Founding Fathers and *Huckleberry Finn*, and the separa-



tion of Church and State.

There is an alternative. It's even a legal and simple course of action, though uncommon. The overriding principle here, partly embodied as I have shown in statutory law, is local control. Each school sets its own curriculum, coordinates it with other schools' curriculums in the district as directed by the superintendent, and bases it on *Vermont's Framework of Standards*. Nothing says a school or a district has to draft and write its curriculum document from scratch. And right here my Vermont district displays a certain timidity and conformity in regularly revising its own curriculum guidelines. The existing documents are prevented from providing a specific grade-by-grade content by their arrangement into three clumps of four grades each. As can be seen from the excerpts presented on page 32, the very layout of these documents precludes a sequential curriculum.

The alternative would set aside existing Vermont curriculum documents. My district can examine and evaluate and finally select one from among a number of independent, off-the-shelf curricula now available, both public and proprietary. The New York State Board of Regents, the International Baccalaureate, Success for All, the Edison Project, the Core Knowledge Sequence, Direct Instruction, America's Choice—all of these programs make differing claims. Having spent much time in the past three years scrutinizing these programs and their curricula, I can say that each of these may offer schools something very useful. But, I have found only one curriculum that moves grade-by-grade (in this case K-8), that uses simple lists of specific content, does not prescribe teaching methods (which can be done locally), is cross-referenced, and that turns out to be informative and even a pleasure to read. The *Core Knowledge Sequence* (now in its third edition), prepared and published by the Core Knowledge Foundation in Charlottesville, Va., accomplishes all this in a no-frills 200-page booklet currently being used by nearly 500 K-8 schools. (See the excerpt in the box on p. 33.) The moving spirit here is the dedicated teacher-scholar E.D. Hirsch, Jr.<sup>2</sup> Everyone concerned about what is being taught in our public schools should examine the *Core Knowledge Sequence*. The considered selection of such a curriculum by my district would represent the full and proper exercise of local control and a means of coordinating the preparation of students in the five elementary schools feeding Mt. Abe.

For some schools and for some teachers, so specific a program of study would represent a fundamental change, almost a conversion, and would have to be carefully implemented. With the help of the Core Knowledge Foundation School Department, hundreds of schools have made the transition. For the most part, teachers, students, parents, and administrators have been satisfied with the results. In a school setting it helps enormously when all parties can find out easily just what is being taught in any course and how the sequence fits together to cover the ground.<sup>3</sup>

Vermont has offered its school districts and schools the opportunity to choose the best off-the-shelf curriculum. I'm not a paid lobbyist. I merely hope to demonstrate to my district with its six schools and seven boards, and to anyone concerned with school curriculum, that the *Core Knowledge*

Dewey now concludes that he wants his Laboratory School to be simultaneously child-centered and curriculum-centered. He is not proposing a compromise: He is promoting two complementary viewpoints.



*Sequence* embodies the dynamic balance that Dewey discovered while running the Laboratory School—the balance between the developing child and the mature curriculum.

And just think, students in my district and in other districts might learn to understand references to my proposal as “quixotic.” For in the *Core Knowledge Sequence* that I am championing, episodes from Cervantes’ novel appear prominently in the fifth-grade English curriculum. □

## Notes

<sup>1</sup> See John Dewey, *Dewey on Education: Selections*, edited and with an introduction by Martin S. Dworkin. Teachers College, Columbia University, 1959, pp. 20, 25, 27.

<sup>2</sup> For an article describing E.D. Hirsch’s general approach, see his “The Primal Scene in Education,” *The New York Review*, March 2, 1989.

<sup>3</sup> Full information on the activities and publications (including videos) of the Core Knowledge Foundation is available at its Web site: [www.coreknowledge.org](http://www.coreknowledge.org).



# The “AYP” Blues

*Low-Achieving Schools Will Fail—but They’re Not the Only Ones*

By Nancy Kober

Each year, states release lists of schools and school districts that have not made “adequate yearly progress” (AYP) in raising student achievement. Schools and districts appear on these lists because they have fallen short of one or more of the annual test score targets and/or other performance benchmarks set by states to comply with the federal No Child Left Behind Act (NCLB).

This process of monitoring and reporting on adequate yearly progress is a central concept of NCLB. It is meant to highlight schools and districts that aren’t performing as well as they should and to stimulate actions that will improve teaching and learning. To make AYP, schools and districts must:

- Ensure that a state-determined percentage of students in every major subgroup—including major racial and ethnic minority groups as well as low-income students, disabled students, and English language learners—scores at the “proficient” level on state achievement tests in reading and mathematics;
- Test at least 95 percent of the students in each subgroup and overall;
- Meet at least one other state-determined academic indicator—graduation rates for high schools and attendance rates or another indicator for elementary and middle schools.

The percentage of students expected to score at the proficient level must go up over time until it reaches a goal of 100 percent of students achieving proficiency by 2014. The grade levels tested for AYP also increase over time.

Schools that do not make AYP for one year are placed on a

*Nancy Kober is a consultant to the Center on Education Policy and co-author and editor of the Center’s annual reports on NCLB. This article is an adaptation of the September 2004 issue of CEP’s Test Talk for Leaders. The full version is available in CEP’s Web site at [www.cep-dc.org/testing/](http://www.cep-dc.org/testing/).*



“watch list” and must improve their performance the next school year to avoid future consequences. Schools that receive funds under the federal Title I program face increasingly serious sanctions if they fail to make AYP for multiple years. After two consecutive years of falling short, these schools are identified

for improvement. They must develop a specific plan to boost student achievement, and their school district must reserve a portion of its Title I money to finance transportation for students who choose to transfer to a higher-performing public school. After three or more consecutive years of missing the mark, their district must use a portion of its Title I funds to pay for tutoring services, and Title I schools are subject to stiffer sanctions, ending with replacement of administrative and/or teaching staff, restructuring of school governance, or takeover by a private management company or the state.

For three years, the Center on Education Policy (CEP) has been conducting a comprehensive national study of federal, state, and local implementation of the No Child Left Behind Act. This work has shed light on various reasons why a school or school district may fail to make AYP.<sup>1</sup> Simply put, low-achieving schools will fail to make AYP, but not all schools that fail to make AYP are performing poorly.

## Why Could a School Fail to Make AYP?

Here are some of the most important reasons why schools miss AYP targets.<sup>2</sup>

### 1. Some schools are doing a poor job of educating students.

The most obvious reason why some schools do not make

<sup>1</sup> As this article goes to press, the U.S. Department of Education has announced that states will be given greater flexibility, so some of the issues presented here may be addressed by changing regulations.

<sup>2</sup> For a more comprehensive discussion of reasons why schools fail to make AYP, see [www.cep-dc.org/pubs/TestTalk/CEPTestTalk3.pdf](http://www.cep-dc.org/pubs/TestTalk/CEPTestTalk3.pdf).



AYP is because they are not adequately educating students. Often these schools fall short in several areas. They may have low test scores overall, as well as for specific subgroups of students, and they may also have low graduation rates or poor attendance rates. These are the kinds of schools that NCLB accountability was intended to highlight. For these schools, being identified for improvement can bring technical assistance, additional resources, and other supports to help them do better. It can also mean a loss of enrollments, if parents take advantage of the NCLB choice option, and a diversion of school district Title I funds to pay for student transportation to choice schools or off-site private tutoring services.

### **2. Schools with diverse enrollments often have a more difficult time making AYP than small or relatively homogenous schools because they must meet performance targets for more subgroups.**

Even if a school's overall test scores exceed state AYP targets, the school will fail to make AYP if just one student subgroup misses the mark. Consequently, more diverse schools and districts have a tougher time making AYP than less diverse ones because they have more subgroups large enough to count for AYP (for more information on minimum subgroup sizes, see item three). CEP's study (2005) found that significantly more urban and very large districts had schools that did not make AYP for one or more years than did suburban, rural, or small districts. This is partly because urban and very large districts tend to have more subgroups that count for AYP.

Schools with subgroups may not make AYP because they aren't doing a good job educating students in certain subgroups. They may also be making some progress with their subgroups, but not enough progress to meet the AYP target (for more on this, see item four). Or, they may not make AYP simply because each additional subgroup creates an additional chance for test score fluctuations (see "Accountability 101" on p. 38).

### **3. Whether a school makes AYP depends on which state it is located in.**

The proportion of schools not making AYP varies enormously from state to state. Based on 2003-04 testing, less than five percent of the public schools in Wisconsin did not make AYP, compared with about 76 percent in Florida (Wisconsin Department of Public Instruction, 2004; Florida Department of Education, 2004). At least some of this disparity is the result of different state policies for implementing AYP, rather than differences in student learning. Because of these differences, comparisons of state AYP lists are not a meaningful indicator of the quality of schools in one state versus another.

The No Child Left Behind Act lets states make several key decisions about their accountability systems. For example, each state:

- Determines its own curriculum standards;
- Develops or chooses its own tests to measure progress toward its standards;

- Decides where to set the cut scores on state tests to define "proficient" performance for AYP;

- Sets targets beyond the starting point<sup>3</sup> for the percentage of students that must score at the proficient level each year, as long as these targets rise periodically and are set at 100 percent proficient by 2014;

- Decides how large subgroups must be in order to be counted for AYP;

- Decides whether to use statistical methods like confidence intervals to compensate for test score fluctuations stemming from factors unrelated to learning (see "Accountability 101" on p. 38);

- Selects and sets standards for determining high school graduation rates and, in grades three through eight, chooses another indicator of performance, such as attendance.

Together, these decisions affect how easy or hard it is for schools to make AYP.

One area of state difference relates to test difficulty, test cut scores, and targets for the percentage of students reaching proficiency. Some states use less rigorous tests while others administer harder tests. Some states have lower cut scores for proficient performance while others have higher ones. In 2004-05, Tennessee fourth-graders had to correctly answer 36 percent of the items on the state test to achieve the proficient level in language arts, while New York fourth-graders had to get 67 percent right (Tennessee Department of Education; New York State Department of Education, 2005). Some states have set steadily rising targets for the percentage of students that must score at proficient levels on the way to 100 percent proficiency, while other states have "back-loaded" their trajectories, setting relatively low targets in the early years and calling for faster progress in later years. By the law's formula, Oregon, for example, had a starting target of 39 percent proficient in math in 2002-03 and chose to increase it to 49 percent in 2004-05 (Oregon Department of Education). Virginia had a starting target of 58.4 percent proficient in math in 2002-03 and chose to increase it to 70 percent proficient in 2004-05 (Virginia Board of Education). Although some states with higher cut scores have lower proficiency targets and vice versa, this is not always the case; the interactions of test difficulty, cut scores, and proficiency targets do not seem to follow a clear pattern. Studies have shown, however, that students with exactly the same knowledge and skills would miss the proficiency target in some states and easily surpass it in others (Helderman and Mui, 2003).

State differences in minimum subgroup size also influence whether a school makes AYP. In New Mexico, a subgroup must include at least 25 students to be separately counted for AYP purposes, while in Virginia the minimum size is 50 students (New Mexico State Department of Education; Virginia Board of Education). During the past year, several states increased their minimum subgroup size, in ef-

<sup>3</sup> Each state's starting point is calculated through a formula in the NCLB law.



A school in a poor neighborhood that steadily boosted its achievement from 25 percent proficient in 2004-05 to 40 percent in 2007-08 could fail to make AYP.



school score below proficient one year, and 67.5 percent score below proficient the next year, and the graduation rate increases sufficiently, the school will make AYP. Even so, safe harbor is a challenge because each major subgroup that misses the fixed AYP targets must separately meet the 10 percent improvement criteria to qualify for safe harbor.

Exceptions also exist in the limited number of states that use a performance index as part of their NCLB accountability systems. These states, which include Massachusetts, Minnesota, New York, and a few others, give schools or districts partial credit toward AYP for partially proficient students (American Federation of Teachers, n.d.). Minnesota, for example, assigns half an index point for each student reaching the partially proficient level and one full point for each student reaching the proficient level. These indices may help some improving schools make AYP that otherwise would have failed. Only a handful of states have sought approval for index systems from the U.S. Department of Education.

fect making it easier for schools to meet AYP targets because fewer schools have subgroups large enough to count (CEP, 2005).

#### 4. A school could raise achievement for struggling students but still not make AYP if these students score below the proficient level.

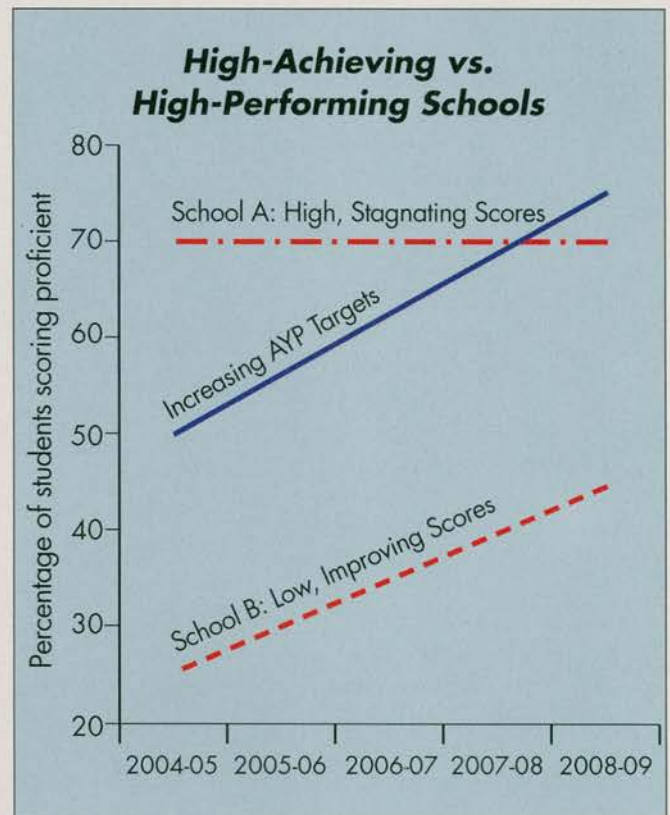
Rather than tracking the progress of individual students, AYP is generally based on the percentage of students reaching a fixed proficiency target. Students who score far below the proficient level have much more ground to cover than students who are closer to proficiency. But with few exceptions, schools do not receive AYP credit for improving the achievement of these lowest-performing students. Nor do they get credit for raising the achievement of high-performing students who have already reached proficiency.

Consider two hypothetical schools (portrayed in the figure on the right) in a state with an AYP target of 50 percent proficient in math for 2004-05, rising to 70 percent proficient in 2007-08. A school in which 70 percent of students already score proficient could post no gains (or could even show declines) for a few years and still make AYP through 2007-08. But a school in a poor neighborhood that steadily boosted its achievement from 25 percent proficient in 2004-2005 to 40 percent in 2007-08 would fail to make AYP. In short, the first school could stagnate, while students and teachers in the second school could easily become discouraged because their gains don't count. The same situation could arise with high-performing and low-performing subgroups.

NCLB's "safe harbor" provision offers an exception to this principle by giving credit for significant improvements below the proficient level. If a school does not meet state AYP targets, but it reduces by 10 percent the share of students scoring below proficient and also makes progress on the state's other academic indicator, then it will make AYP under safe harbor. For example, if 75 percent of the students in a high

#### 5. AYP hurdles become more numerous and higher over time.

By 2006, students must be tested yearly in reading and math in grades 3 through 8 and once during grades 10 to 12. By 2007-08, students must be tested in science, too, although states can decide whether to include science results in their AYP calculations. Many states are still phasing in all the





NCLB-required tests. In 2004-05, about half the states were administering reading and math tests in all the required grades, and 23 states were administering science tests (Olson, 2004). The expansion of testing to additional grades in many states and to another subject in some states could create more opportunities for schools to miss AYP targets.

Even more significantly, the AYP hurdles get higher over time as states move closer to 100 percent proficiency. Given the normal fluctuations that occur in yearly test scores (see "Accountability 101" below), it is unrealistic to expect every school and district to post steady achievement gains year after year. In addition, several states and districts participating in CEP's study (2005) thought that the requirement of 100 percent proficiency was unrealistic, particularly for the

subgroups of students with disabilities and English language learners. Whether or not schools surmount the current set of AYP hurdles, the bar continues to rise. An analysis of AYP trends in Connecticut (which is consistently one of the highest-scoring states on the National Assessment of Educational Progress) concluded that as expectations are raised and more grades are tested, 90 percent of the elementary and middle schools in the state will fail to meet AYP targets by 2014 (Frahm, 2004). Other states have projected similar outcomes.

**6. AYP is determined largely by looking at whether the students in the tested grades of a school reach fixed proficiency targets each year. Changes in the composition of**

## Accountability 101: Tests Are Blunt Instruments

Large-scale tests like those used for NCLB have advantages over less standardized forms of measurement. They can provide results that are more consistent and useful for comparisons than those from assessments based on individual judgment. They can also produce extensive information about student performance at lower costs and with less testing time than many other forms of assessment. Because large-scale tests are developed in a scientific manner and report results in numbers, many people assume they are very precise. But even well-designed tests have limitations that should be considered by users of AYP and other test-based data.

As testing experts often note, a test score is more like an estimate than an exact measurement. If a student took the same test on consecutive days without studying in between, the student's scores may still vary due to factors unrelated to learning, such as the sample of questions on the particular test version, the student's physical condition or state of mind, lucky guesses, or errors in recording answers. Aggregate scores for a group of students—whether a school, classroom, or NCLB-defined subgroup—may also

fluctuate due to factors unrelated to teaching and learning, such as yearly changes in the test-taking population.

Here are some aspects of testing that could produce these types of test score fluctuations:

**a. A test is a sample of all possible questions that could be asked about a subject.** The questions on a test are merely a sample of the vast store of knowledge and skills in a subject like math. A test that lasts a few hours can't possibly address all the important topics, concepts, or math skills that students are expected to learn during the school year.

Test developers try to minimize the impact of this form of sampling variation by selecting questions that cover a representative sample of important knowledge and skills in the subject being tested. They also try to ensure that different versions of the same test (developed for security reasons and to limit teaching to the test) are parallel in content and difficulty. Still, there will always be students who would have scored higher if a particular test version had included a different sample of questions that they happened to know well.

**b. A test administration is a sample of a student's behavior at a single point in time.** On any given day, a variety of external factors—a headache, an argument with a parent that morning, a jackhammer or barking dog outside the school—could negatively af-

fect a student's performance. If the test had been given at another time, the student might have scored higher.

When student scores are combined across a large enough group, fluctuations in individual test scores due to sampling variations in test questions and external conditions tend to offset each other. For example, the uncharacteristically low performance of a student with a headache on test day might be offset by the unexpectedly high score of another student who felt rested and confident and made a few lucky guesses. Since AYP is calculated by looking at the percentages of students scoring at the state-set "proficient" level on state tests, the scores of a few students can mean the difference between making or not making AYP, especially when it comes to subgroups, which include fewer students.

**c. Yearly changes in the test-taking population can produce fluctuations in aggregate test scores.** As any teacher can attest, each year's group of students represents a unique mix of economic, linguistic, and racial/ethnic backgrounds and different capabilities, personalities, and behavior. Countless factors can change the composition of test-takers from year to year in ways that affect aggregate scores. Family income, for example, is a strong predictor of student test scores, so the loss of a major manufacturer could increase poverty and lead to lower aggregate test scores for a school in that community.

*This sidebar is an adaptation of the October 2002 issue of Test Talk for Leaders published by the Center on Education Policy. The original is available on CEP's Web site at [www.cep-dc.org/testing/](http://www.cep-dc.org/testing/).*



the test-taking population could affect the ability of one year's group to meet the fixed targets, especially for smaller schools or subgroups.

The composition of a class, school, or subgroup often varies from year to year in ways that can affect test scores. Students move in and out of the district. One year's group could have an unusually high number of English language learners due to an influx of immigrant children. The loss of a major manufacturer could mean higher poverty and more family disruption. These year-to-year changes can make AYP determinations unreliable, especially for smaller schools or subgroups or in states that do not apply statistical tools such as confidence intervals (see "Accountability 101" below).

Or a school could experience an influx of immigrants, adding more English language learners to the test-taking population. This year's third grade (a tested grade in many states) could have a higher share of students with severe disabilities than last year's third grade. A cluster of students with behavioral problems could create an unusually disruptive classroom environment. An exodus of high-achieving students from neighborhood schools to private or charter schools or construction of an upscale housing development in the neighborhood could change the test-taking group in meaningful ways.

If the number of test-takers is large, these types of yearly changes may have little effect. But in a relatively small group (fewer than 100 students, according to Haney, 2002), annual changes in group composition can produce wider fluctuations because each student's score has a greater impact on the aggregate. With the average elementary school containing only 68 students per grade, score instability is not unusual. It is also common among schools with high mobility or very diverse enrollments.

These fluctuations matter because under NCLB, aggregate scores of students in tested grades and subgroups are used to make judgments about the effectiveness of the entire school. When a school fails to make AYP, people generally don't consider whether the students tested that year are truly

representative of the broader universe of students served by that school across the years, and as a result, they don't question what the test scores truly say about the school's effectiveness.

**d. Most, though not all, states use confidence intervals to compensate for test score variations.** Recognizing that tests are not precise instruments, some states are using a statistical tool called "confidence intervals" to make allowances for score fluctuations unrelated to changes in achievement. Somewhat like the margin of error in a public opinion poll, a confidence interval creates a window around the state AYP target of plus or minus a few points. Test results that fall slightly below the target but within the window are counted as having met the target, so confidence intervals make it somewhat less likely that a school or subgroup will fail to make AYP due to chance fluctuations. The size of the window is determined by two factors: the number of students tested and the degree of confidence that test administrators wish to have in the accuracy of the results. The smaller the group of students tested, the wider the window.

Imagine that 40 percent of the students in a school score at the proficient level in math. Using a 95 percent confidence interval, test administrators can be 95 percent certain that the true achievement of the school falls within a range of 35 to 45 percent proficient. If the AYP target is 42 percent profi-

## What Does This Mean for Schools and Districts?

All of the factors mentioned above, plus others not cited here, create a long, difficult course with many hurdles that schools and districts must traverse to make AYP. Moreover, the sanctions are the same for schools that miss by a little or miss by a lot. For this reason, some educators, researchers, and policymakers have concluded that the current AYP requirements are unrealistic or unfair.

Nonetheless, the NCLB requirements have produced some positive outcomes. They have pushed schools and communities to pay more attention to children and sub-

*(Continued on page 48)*

cient, then the school makes AYP because its score falls within the window. If test administrators want to be 99 percent confident that true achievement falls within the window, then the window would have to be much wider.

The majority of states currently use confidence intervals for various AYP decisions, more than in earlier years (Center on Education Policy, 2005). In states that do not use confidence intervals, one school could be labeled as low-performing while another is deemed adequate, even though there's no meaningful difference between their aggregate scores.

Sampling variation in test questions and external conditions and changes in group composition can produce yearly fluctuations in test scores unrelated to educational effectiveness. A school's test score trends across several years can provide a trustworthy indicator of student achievement, but a single year's scores may or may not be a good indicator of the quality of teaching and learning in a school.

—N.K.

## References

- Center on Education Policy. (2005). *From the capital to the classroom: Year 3 of the No Child Left Behind Act*. Washington, D.C.: Author.
- Haney, W. (July 10, 2002). "Ensuring failure: How a state's achievement test may be designed to do just that." *Education Week*.



# Ideas to Consider: How to Make AYP Work Better for Students and Their Schools

*There is a growing consensus that, as described in the previous article, the formula for determining whether a school has made adequate yearly progress (AYP) is deeply flawed. These flaws can produce misjudgments about a school's effectiveness and then, under NCLB, trigger consequences that impede, rather than support, school improvement. (For example, a school that is improving could be shut down or turned into a charter.) Here are ideas culled from the worlds of policy and research on a few ways to overcome some of AYP's main problems.*

—EDITORS

## 1. Establish ambitious, but attainable, goals for how much progress schools should make.

Goals for school progress mandated by an accountability system should be ambitious, but they also should be realistically obtainable with sufficient effort. At the very least, there needs to be an existence proof. That is, there should be evidence that the goal set for all schools does not exceed what has been accomplished by those schools that have made the greatest progress. For example, you could identify the 10-20 percent of high-poverty schools that had the greatest average rate of increase over three to five years and establish that rate of improvement as a target to shoot for. That would be a great challenge to the vast majority of schools, but it might be a target that is within reach with sufficient effort. Such evidence-based performance goals are necessary under AYP—and would be necessary in a value-added system as well.

—ROBERT LINN,  
*Distinguished Professor of Education,  
University of Colorado at Boulder*

*The comment by Chester Finn and Frederick Hess is drawn from "On leaving no child behind" in the Fall 2004 issue of Public Interest. Michael Barber's comment is drawn from a lecture he gave at Boston University that is published in the current issue of the Journal of Education.*



## 2. Evaluate schools based on students' progress, not their status, using a genuine "value-added" system.

*Many critics of No Child Left Behind's AYP provisions point to two related problems with the current system that teachers immediately understand: First, how well kids do in school depends partly on things that happen to them outside of school, especially at home. Second, schools shouldn't be judged simply on their students' absolute test scores (since so much of that depends on out-of-school factors); they should be judged by how much progress they help their students make—their "value added." Here's the view of one statistician.*

Adequate yearly progress (according to NCLB) is primarily determined by the proportion of students who are "proficient" on the state's accountability test. It depends on the students' achievement status at a point in time. One concern with using achievement status for accountability is that it depends largely on factors unrelated to school quality (such as what a child already knows when he enters kindergarten, what he learns during the summer, etc.) and is highly correlated with family attributes such as wealth and parents' level of education. Critics of NCLB and AYP often complain that by focusing on achievement status, the law evaluates school quality based on factors that are unrelated to schooling.

A potential alternative that is gaining

popularity among researchers, educators, and policymakers is to evaluate schools based on individual students' growth in achievement. This is an intuitively appealing idea because schools contribute to student learning and learning should be visible as growth in achievement, regardless of the students' initial status. Studies find that although growth can also be related to family and neighborhood attributes, it is much less related than achievement status is. Therefore, measuring growth is seen as a way to focus on schooling inputs to education and is often referred to as "value-added" assessment. Value-added assessment systems follow individual students' achievement over several years. The term should not be confused with other measures that compare successive cohorts of students (e.g., this year's fourth-graders to last year's), but do not report on how the same students do on tests over several years.

Proponents of value-added assessments argue that they more accurately portray the actual quality of schools and point to the examples of schools with low achievement status but high levels of growth or schools with high levels of achievement but low growth to argue that value-added measures can truly identify high or low performing schools rather than merely identifying schools that serve high or low achieving populations. (For a hypothetical example, see the figure on p. 37.)

Value-added assessments clearly have their benefits, but there are challenges to overcome before states build them into their accountability systems. First, although they are an improvement over achievement status measures, value-added assessments might not provide accurate measures of schools' contributions to learning, and methods to account for such errors are being developed. Second, the statistical procedures used—such as adjustments to minimize the effects of family and neighborhood attributes—are complex and lack the transparency that many people consider



desirable in accountability systems. Third, measuring growth requires linking individual students' scores across grades and places great demands on the tests to ensure that growth is due to student learning, not differences in the tests. Many state tests and data systems do not currently meet these rigorous requirements. Fourth, in order for value-added assessments to contribute to an evaluation of school quality, standards for adequate growth need to be set. However, meeting a standard of growth will not ensure meeting a standard of proficiency. As I heard one policymaker say, "How do you explain to parents that their children's schools provided good value-added growth for the last 12 years, but their children still failed to pass the state's high school exit exam?" Possibly different standards of growth will be required for different schools. For these (and other) reasons, any move to value-added accountability systems should be paired with well-designed evaluations to give us a better understanding of the best uses of the measures and their potential limitations.

—DANIEL MCCAFFREY,  
*Statistician, The RAND Corporation*

### 3. Require statistical safeguards. But don't use them to take the focus off of traditionally low-achieving groups.

As discussed in "Accountability 101" (p. 38), there are many reasons, unrelated to teaching and learning, why a school's test scores might not accurately reflect the school's quality. To help assure that these inaccuracies don't lead schools to be misidentified, it's critical that states make use of such statistical safeguards as confidence intervals and multi-year averaging; not all states currently do so. These safeguards are critical to any accountability system, not just NCLB's, and are even more important when accountability is based on achievement growth.

The need for statistical safeguards is especially important when dealing with small groups, such as the subgroups identified by NCLB. Most states try to address this by establishing a minimum subgroup size—meaning that if a subgroup is too small, its progress won't

count for AYP determination. Unfortunately, not only is this a rather crude statistical safeguard, but it also undermines NCLB's promise to "leave no child behind" by allowing states to avoid counting the test scores of small groups of disadvantaged children. Confidence intervals do a better job of ensuring statistical reliability by taking into account the statistical "margin of error" in the AYP calculation in the same way that public opinion polls include a margin of error. Because the margin of error widens as subgroups get smaller, large minimum subgroup sizes are not needed—and more subgroups can count for AYP determinations.

Several states are now "gaming" AYP; by establishing excessively large minimum subgroup sizes, especially for the special education subgroup, states seem to be focused primarily on reducing the number of subgroups counted in AYP determinations rather than on fairness and accuracy.

—HOWARD NELSON,  
*Senior Researcher, American Federation of Teachers*

### 4. Complement the evidence provided by test scores with on-the-ground observations.

*In Britain, where greater school accountability has also been introduced, an inspection system, in which trained inspectors observe schools, is used to complement the data produced by test scores. An aide to Britain's prime minister explains one of the reasons why.*

Inspection enables a much more refined approach to dealing with school failure. Intervention in schools that are seriously underperforming—enabled by the development of accountability—has been hugely beneficial, but where the system depends purely on test results to determine school failure or success, it risks being far too crude. Our interventions in failing schools are driven by the inspection system. Where a team of inspectors judges a school to be failing ("in need of special measures," as the legal euphemism puts it), a second team of inspectors follows up shortly afterwards to corroborate the judgment. This process enables real analysis—not just of whether performance is poor, but also why. In addition, it enables the

system to identify and tackle failure even where it is masked by temporarily reasonable test results.

Once a school is in special measures, the inspectors return three times a year. Often within a year or 18 months, they are able to give a school a clean bill of health. Our evidence suggests that in these circumstances, the expertise of the inspectors is hugely appreciated. For the principal and staff, these visits are simultaneously both challenging and beneficial. They provide an expert commentary to the school on what is happening. There is feedback on the impact of changes in leadership, standards of attendance and behavior, staff morale, and the systems in place for grading work, dealing with pupils with special educational needs, and so on. These changes are the lead indicators that point to improvements in test scores in the future. A system depending purely on test scores both for intervening and for deciding whether the intervention has worked has no such subtlety and can sometimes have destructive consequences.

—MICHAEL BARBER,  
*Head of the Prime Minister's Delivery Unit,  
United Kingdom*

### 5. Distinguish among schools that are progressing substantially, schools that need to improve, and schools that desperately need to improve.


NCLB should replace its all-or-nothing AYP calculation with a more flexible approach. One might, for example, distinguish among schools that are making progress overall and in 90 percent or more of their demographic subcategories; those that are making progress overall but in less than 90 percent of categories; and those failing to make acceptable overall progress. Such a triage system would reduce the vast number of mostly okay schools that are now being labeled as "needing improvement." It would distinguish between those that are on the verge of succeeding and those that are catastrophically inadequate. It would enable states and districts to focus on repairing the latter.

—CHESTER FINN,  
*President, Thomas B. Fordham Foundation,  
and* FREDERICK HESS,  
*Director, Education Policy Studies,  
American Enterprise Institute*









to walk down and look at what was going on. "But when Tilly explained what she thought was going to happen, I had second thoughts. We ran off the beach as fast as we could and went up to the first floor of the hotel. I dread to think what would have happened if we had stayed. Minutes later the water surged right over the beach and demolished everything in its path. It was terrifying to watch but I am very proud of her. The tragedy on Boxing Day was awful, we are so lucky to be alive."

Mr. Kearney, who teaches Tilly at Oxshott's Danes Hill Prep School, said her quick-witted actions were typical. Mr. Kearney, 56, added:

I'm stunned at the news but so relieved she and her family are safe. Tilly is a very bright, level-headed girl. Nothing illustrates her character more than her brave actions in a terrifying situation.

It is an incredible coincidence that our class was learning about this type of tsunami just two weeks before Christmas.

I showed the class a video

clip of a Hawaiian tsunami disaster during the 1950s. A woman survivor described how the waters dramatically receded and then came hurtling back minutes later.

I distinctly remember telling the students that after the sea was sucked backwards the next five to ten minutes were crucial for people to survive. I drew pictures of a tsunami on the blackboard and all the students copied them into their exercise books. They were also told of the devastating effects when those waves hit shallow water and how people can easily fail to spot the danger when the waters pull back.

Penny, dad Colin, 46, Tilly, and Holly, 7, at first planned to leave Thailand after the disaster. But they responded to pleas from the Thai tourist board and finished their holiday. Penny added: "It was a difficult decision but the thing the locals are dreading is tourists leaving. It will be another tragedy for Thailand if tourists cancel holidays." Craig Smith, general manager of the five-star JW Marriott Hotel where Tilly and family stayed, said, "She's a real heroine. I think it's phenomenal that Tilly's parents and the others on the beach are alive because she studied hard at school."

Tilly's headmaster Robin Parfitt added, "Tilly has a sharpness and wisdom beyond her years. Our motto is 'Strong and Sagacious.' It sums up Tilly pretty well." □



## Standards-Based Reform

(Continued from page 19)

Nonetheless, the goals of standards-based education were much broader and higher than this. If we want to realize the benefits of standards-based education for the full range of students, and if we want our lowest performing students to reach the high standards that were originally the hallmark of the standards movement, accountability that is so heavily tied to poor tests and that doesn't assure that teachers get the support they need to teach to the standards will not get us there.

### III. Where Do We Go from Here?

If we mean to realize the benefits of standards-based education for the full range of students, and if we want all of our students to reach the high achievement levels that were originally the goal of the standards effort, we will have to attend to more than tests and accountability systems. The nation's efforts to truly realize the goals of standards-based education will be frustrated by the incompleteness of the reforms that have been put in place so far. The instructional support system—curriculum, instructional programs, professional development, targeted interventions for struggling students—that now exists in most districts is not strong enough to produce achievement that goes beyond bringing the basics to a larger group of students. And, if we continue to neglect the core of the reforms while pressing forward on accountability, we will engender more and more hostility from the public and educators who want what standards-based reform promised: truly high standards for all. Each of the four key elements of the standards-based education system needs attention.

**Standards.** States need to strengthen the specificity and clarity of their standards so that these standards can adequately play the role of other countries' national curricula. Standards should be clear, specific, and comprehensive enough to serve as the basis for building both good examinations and strong instructional programs. Grade-by-grade standards seem to provide the best guidance. So do standards that specify the kinds of texts to be read, the particular scientific or mathematics concepts to be learned, and detailed and understandable criteria for good writing and other complex skill performances. The danger in a move to specificity—long lists of topical content or mechanical skills to be mastered—has made some educators wary of detailed standards. It is time to take a new look and to find a "Goldilocks" solution—a workable middle ground between too much and too little detail in standards. In this difficult process, states may find it useful to borrow from one another or from existing published syllabi and standards, or to join consortia that are developing shared standards. An example is Achieve's Mathematics Achievement Partnership, a group of nine states working together to raise expectations and improve student performance in middle- and high-school mathematics. As part of the initiative, Achieve has published a framework for what American students need to know in mathematics in the middle grades (Achieve, 2001).

---

Because the stakes are high, the incentives to match teaching to tests instead of standards are almost irresistible. And if we don't sharpen standards and assess what we really mean by them, the nation is likely to wake up in a few years to find that it has created a "fool's gold" system.

---

**Assessments.** There is much to do, as well, on the testing and assessment front. Assessments play a dual role in a standards-based system. They are instruments for monitoring and accountability and, at the same time, they inevitably model and guide instruction. The higher the stakes, the more educators will feel pressed to teach to the tests. Therefore, the higher the stakes, the more important it is that assessments guide educators, and students, toward the kind of learning we truly want. We have seen that most of today's state tests are not well aligned to standards and they are most likely to leave out the most intellectually challenging aspects of the standards. Yet, it is the tests rather than the standards that claim educational attention.

To recapture the intent of the standards-based system, most state assessments need to be redesigned so that they guide teaching in the direction really intended by the standards. This will probably require adding substantial numbers of tasks that require open-ended and constructed responses, as is the practice virtually everywhere else in the world. There is no mystery about how to do this in ways that meet technical standards of measurement. But there is no doubt that standards-referenced assessments that include substantial numbers of open-ended and constructed-response tasks will be costly. Substantial assistance from the federal government is likely to be needed by all but the largest states. Dollars granted to states or consortia for this essential work will help.

**Instructional programs and professional development.** With standards and assessments still needing substantial work, it is perhaps not surprising that instructional programs and professional development geared to standards are barely out of their infancy. Here, too, we will need "Goldilocks" solutions that provide guidance that is detailed enough so that teachers don't each have to invent their own program, while leaving enough room for adaptation to students. Recent research has made it clear that professional development works best when it is tied directly to the program that teachers are using with their students. Programs that re-



quire teachers to follow word-by-word scripts and extremely prescriptive time schedules are unlikely to engage the best minds and the most committed educators for long. But leaving teachers to guess at what are the best ways to teach does not work either. Again, the task is to find the right balance. But even the best instructional programs will fail with some students; the structure and resources must be available to provide these students with intensive interventions.

Where districts develop their own instructional and professional development systems, “buy-in” may be greater. But only the largest school systems usually have the resources for full program development. States may need to provide more tools and direct assistance than they do now, as well as more financial resources. And there is a key role for the federal government in supporting the development and testing of the kinds of research-based instructional systems that we have referred to as “designed programs.”

**Accountability.** Forms of accountability that keep the education system focused on important academic achievement goals and on equity—providing a high-quality education to all of our students—are essential. As we have noted, many aspects of the current NCLB accountability requirements need to be adjusted. For example, ways need to be found to measure and reward achievement growth, thus taking into realistic account our schools’ different starting points. And we need to reconsider some of the accountability requirements for special education students and English language learners. Thoughtful individuals in the states, the federal government, and the research community are at work on these issues, and we remain confident that the accountability provisions of NCLB can be adjusted as we learn more about how it actually works.

However finely tuned the accountability rules, however, they cannot have their intended effect on the quality and accessibility of education unless the first three components of the standards-based system are brought up to par. The accountability aspect of the program is, if anything, running dangerously ahead of the system as a whole. Because the stakes are high, the incentives to match teaching to tests instead of standards are almost irresistible. And if we don’t sharpen standards and assess what we really mean by them, the nation is likely to wake up in a few years to find that it has created a “fool’s gold” system. We will have more and more of the least valuable coin of the realm; while the high levels of achievement we meant to create will increasingly elude us.

**M**uch has been accomplished since the National Governors Association summit that put standards on the front burner. But, increasing student achievement beyond a relatively low standard will be nearly impossible unless we create the coherent whole that inspired the standards movement 15 years ago.

The original vision of standards-based education, we think, was the right one. And some notable progress has been made: Standards are now in place, although some will need substantial revision before they can adequately guide educators toward the intended high expectations for learning. Accountability has

produced unprecedented attention to the very students it had been easy to ignore, or to set low expectations for, in the past. But standards and accountability are only the outer shell of the standards vision. The core of the reforms—aligned, high-quality assessments, instructional programs, and professional learning opportunities—have yet to be realized. □

## References

- Achieve, Inc. (2002). *Staying on Course: Standards-Based Reform in America's Schools: Progress and Prospects*. Washington, D.C.
- Achieve, Inc. (2001). *Foundations for Success: Mathematics for the Middle Grades*. Washington, D.C.: Mathematics Achievement Partnership.
- American Federation of Teachers (2001). *Making Standards Matter*. Washington, D.C.: American Federation of Teachers.
- Barnett, W.S., (2005). Personal communication.
- Barnett, W.S., Huystedt, J.T., Robin, K.B., Schulman, K.L. (2004). *The State of Preschool: 2004 State Preschool Yearbook*. New Brunswick, N.J.: National Institute for Early Education Research.
- Boyer, E. L. (1983). *High School: A Report on Secondary Education in America*. New York: Harper & Row.
- Bureau of Labor Statistics (2004). *Employment Status of the Civilian Noninstitutional Population, 1940 to date*. Washington, D.C.: U.S. Department of Labor.
- Center on Education Policy (2005). *From the Capital to the Classroom: Year 3 of the No Child Left Behind Act*. Washington, D.C.: Center for Education Policy Report.
- CGCS (2005). *Beating the Odds: A City-By-City Analysis of Student Performance and Achievement Gaps on State Assessments. Results from the 2003-2004 School Year*. Washington, D.C.: Council of the Great City Schools.
- Cobb, P., Wood, T., Yackel, E., Nicholls, J., Wheatley, G., Trigatti, B., and Perlwitz, M. (1991). “Assessment of a problem-centered second-grade mathematics project.” *Journal for Research in Mathematics Education*, 22, 13-29.
- Cohen, D.K. and Hill, H.C. (2001). *Learning policy: When state education reform works*. New Haven, Conn.: Yale University Press.
- Debra P., et al. v. Turlington, et al., 474 F. Supp. 244 (Fla. Dist. 1979).
- Education Week* (2004). *Quality Counts 2004: Count Me In*. January 8, 2004.
- Jennings, J. and Hamilton, M. (April/May 2004). “What’s good about public schools.” *Our Children*.
- Klein, D., Braams, B.J., Parker, T., Quirk, W., Schmid, W., Wilson, W.S., Finn, C.E., Torres, J., Braden, L., and Raimi, R.A. (2005). *The State of State Math Standards 2005*. Washington, D.C.: The Thomas B. Fordham Foundation.
- Mandlawitz, M. Ed. (2005). *Education Budget Alert for Fiscal Year 2006*. Washington, D.C.: Committee for Education Funding.
- McCutchen, D., Abbott, R. D., Green, L. B., Beretvas, N., Cox, S., Potter, N. S., Quiroga, T., Gray, A. L. (2002). “Beginning literacy: Links among teacher knowledge, teacher practice, and student learning.” *Journal of Learning Disabilities* 35(1), 69-86.
- Merriman, D. (2000). *What Accounts for the Growth of State Government Budgets in the 1990s?* Washington, D.C.: The Urban Institute.
- National Center for Education Statistics (2003). “High School Graduates Compared With Population 17 Years of Age, by Sex of Graduates and Control of School: Selected Years, 1969-70 to 2002-03.” *Digest of Education Statistics and Figures 2003*. Washington, D.C.: U.S. Department of Education.
- National Center for Education Statistics (2000). *NAEP 1999 Trends in Academic Progress: Three Decades of Student Performance*. Washington D.C.: U.S. Department of Education.
- National Commission on Excellence in Education (1983). *A Nation at Risk: The Imperative for Educational Reform*. Washington, D.C.: U.S. Government Printing Office.
- National Council of Teachers of Mathematics (1989). *Curriculum and evaluation standards for school mathematics*. Reston, Va.: NCTM.
- Olson, L. (2003). “Standards and Tests, Keeping Them Aligned,” *Research Points*, 1 (1) 2003. Washington, D.C.: American Educational Research Association.
- Olson, Lynn (2002). “Survey Shows State Testing Alters Instructional Practices.” *Education Week*, Volume 21, Issue 32, pp. 14, April 24, 2002.
- Powell, A., Farrar, E., and Cohen, D. (1985). *The Shopping Mall High School: Winners and Losers in the Educational Marketplace*, Boston, Mass.: Houghton Mifflin.
- Porter, A.C. (2002). “Measuring the Content of Instruction: Uses in Research and Practice,” *Educational Researcher*, Vol. 31, No. 7, pp. 3-14.
- Resnick, D.P. and Resnick, L.B. (1985). Standards, curriculum, and performance: A historical and comparative perspective. *Educational Researcher*, 14(4), 5-20.



- Resnick, L.B. and Resnick, D.P. (1990). Tests as standards of achievement in schools. In G.R. Anrig (Ed.), *The uses of standardized tests in American education: Proceedings of the 1989 ETS Invitational Conference* (pp. 63-80). Princeton, N.J.: Educational Testing Service.
- Resnick, L.B. and Resnick, D.P. (1992). Assessing the thinking curriculum: New tools for educational reform. In B.R. Gifford and M.C. O'Connor (Eds.), *Changing assessments: Alternative views of aptitude, achievement, and instruction* (pp. 37-75). Boston: Kluwer.
- Ryan, J. E. (2004). *The perverse incentives of the No Child Left Behind Act*. New York University Law Review, 932-989.
- Sizer, T. (1984). *Horace's Compromise: The Dilemma of The American High School*, Boston, Mass.: Houghton Mifflin.
- Smith, M. and O'Day, J. (1991). "Systemic school reform." In S. Fuhrman and B. Malen (Eds.), *The politics of curriculum and testing*. Philadelphia: Falmer Press.
- Sonnenberg, W.C. (2004). *Federal Support for Education FY 1980 to FY 2003*. U.S. Department of Education. Washington, D.C.: National Center for Education Statistics.
- Stotsky, S. and Finn, C.E. (2005). *The State of State English Standards 2005*. Washington, D.C.: The Thomas B. Fordham Foundation.
- Viadero, D. (1994) "Teaching to the Test." *Education Week*, Vol. 13, Issue 39 (Extra Edition), pp. 21-25. July 13, 1994.
- Grissmer, D., Flanagan, A., Kawata, J., and Williamson, S. (2000). *Improving Student Achievement: What State NAEP Test Scores Tell Us*. Santa Monica, Calif.: RAND Corporation.
- Grissmer, D. and Flanagan, A. (1998). *Exploring Rapid Achievement Gains in North Carolina and Texas: Lessons from the States*. Washington, D.C.: National Education Goals Panel.
- Johnson, J., Duffett, A., Vine, J., and Moye, L. (2003). *Where We Are Now: 12 Things You Need to Know about Public Opinion and Public Schools*. New York: Public Agenda.
- National Center for Education Statistics (2001). *Entering Kindergarten: A Portrait of American Children When They Begin School: Findings from the Condition of Education 2000*. Washington, D.C.: U.S. Department of Education.
- Office of Educational Research and Improvement (January 1994). "What Do Students' Grades Mean? Differences Across Schools." *Office of Research, Education Research Report*. Washington, D.C.: U.S. Department of Education.
- Rebell, M. (2004). *Adequacy Litigations: A New Path to Equity?* New York, N.Y.: ACCESS, a project of the Campaign for Fiscal Equity.
- Roderick, M., Engel, M., Nagaoka, J., et al. (February 2003). *Ending Social Promotion, Results from Summer Bridge*. Chicago, Ill.: Consortium on Chicago School Research.
- Rose, L. and Gallup, A. (September 1998). "The 30th Annual Phi Delta Kappa/Gallup Poll of the Public's Attitudes Toward the Public Schools." *Phi Delta Kappan*.
- Rose, L. and Gallup, A. (September 2002). "The 34th Annual Phi Delta Kappa/Gallup Poll of the Public's Attitudes Toward the Public Schools." *Phi Delta Kappan*.
- Snipes, J., Doolittle F. and Herlihy, C. (2002). *Foundations for Success: Case Studies of How Urban School Systems Improve Student Achievement*. Washington, D.C.: Council of Great City Schools.
- Vock, D. (2004). "Standards push helps lawsuits: Simply filing a suit—or threatening to—can spur reform. Then political organizing kicks in." *Catalyst*, April 2004.
- White, E. (August 13, 1983). "Poll Finds Public Endorsement of School Reforms." *Education Week*.

## Sidebar References (pages 13 and 16)

- Achieve (2004). *Do Graduation Tests Measure Up?* Washington, D.C.: Achieve, Inc.
- Barnett, W.S., (2005). Personal communication.
- Barnett, W.S., Huystedt, J.T., Robin, K.B., Schulman, K.L. (2004). *The State of Preschool: 2004 State Preschool Yearbook*. New Brunswick, N.J.: National Institute for Early Education Research.
- Beaton, A.E., Mullis, I.V.S., Martin, M.O., Gonzalez, E.J., Kelly, D.L., and Smith, T.A. (1996). *Mathematics Achievement in the Middle School Years: IEA's Third International Mathematics and Science Study (TIMSS)*. Boston, Mass.: Center for the Study of Testing, Evaluation, and Educational Policy.
- Center on Education Policy (2005). *From the Capital to the Classroom: Year 3 of the No Child Left Behind Act*. Washington, D.C.: Center for Education Policy Report.



### Earn a Graduate Degree in Educational Administration on Your Own Time

St. John's University's Graduate Division of The School of Education offers two comprehensive online degrees designed for certified educators who are ready to take the next step in their professional development.

St. John's distance learning degrees offer:

- Flexibility – Manage course work around your family and professional commitments
- Convenience – Study when it's best for you
- Active and engaging online discussions with classmates and professors
- The same outstanding faculty members who teach campus classes
- Designated technical support

Ranked among the top 10% of graduate schools of education in the nation by *U.S. News and World Report*, our online graduate degrees are accredited by Middle States and by the New York State Department of Education.

#### Ready to learn more?

(718) 990-2304  
ronaynek@stjohns.edu  
www.stjohns.edu/distancelearning

54888

## REDEFINING LEADERSHIP



**AZUSA PACIFIC UNIVERSITY**  
ED.D. in EDUCATIONAL LEADERSHIP: K-12

- Focused on Diverse Learners
- Strengths-Based
- Outcomes-Driven
- Cohort Model
- Enhanced with Online Learning
- Real-World Research

*With input from leading national educators, we redesigned our Doctor of Education in Educational Leadership (K-12) Program.*

**Local and national cohorts forming now!**

- CALL** (626) 815-5374
- CLICK** [www.apu.edu/education/k-12](http://www.apu.edu/education/k-12)
- EMAIL** [cjquinn@apu.edu](mailto:cjquinn@apu.edu)
- VISIT** Our Azusa campus, located in L.A. County



For the fourth consecutive year, Azusa Pacific University secured a place in this year's Top 25 of *U.S. News & World Report's* America's Best Colleges Western Universities—Master's classification, ranking 20<sup>th</sup> in a two-way tie from among 124 western colleges and universities.



0305



## School Improvement

(Continued from page 27)

of support and resources schools need to improve. Some of this knowledge can also be found in organizations with staff knowledgeable about school improvement, such as America's Choice, Core Knowledge, and Success For All.

But, finding, organizing, and deploying this expertise is going to require a kind of work that most educators aren't yet very good at and that most policymakers don't know anything about. It will require measures of instructional improvement and performance that are much closer to the ground than the state assessments that are the basis of accountability systems. It will require the creation of systems to find people with expertise in subject matter, instructional practice, and improvement, and getting them into the schools where they are needed. And it will require a new generation of people who are knowledgeable enough about instructional practice to be useful, but who are also interested in broader questions of designing and running support systems.

In short, it will require a lot of hard work in schools and a lot of investment by all of us in the infrastructure of capacity-building. But without this, improvements in school achievement will be small because we won't have provided what Thomas Schelling knew was necessary: We won't have provided people with the knowledge necessary to get the job done. □

## Endnotes

- <sup>1</sup> NCLB does not use the term "failing"; it classifies schools as "in need of improvement." But the law treats the schools as though they are failing, and that is the term I'll use in this article.
- <sup>2</sup> All schools are identified with pseudonyms under assurances of confidentiality.
- <sup>3</sup> See: Elmore, R., Ablemann, C., Even, J., Kenyon, S., and Marshall, J. (2004). "When Accountability Knocks, Will Anyone Answer?" in *School Reform from the Inside Out: Policy, Practice, and Performance*. Cambridge: Harvard Education Press. Also, Carnoy, M., Elmore, R., and Siskin, L. Eds. (2003). *The New Accountability: High Schools and High Stakes Testing*. London: Falmer Press.
- <sup>4</sup> Elementary teachers, I have found, are usually much more knowledgeable and discriminating on this issue. I recently showed a videotape of a high school writing lesson to a mixed group of teachers and administrators from elementary, middle, and high schools. The tape showed a white teacher who had strong skills for engaging his largely minority students in playful and pleasant interactions in class. The work the students produced was, however, obviously very low level, and the teacher's expectations, revealed in his teaching and in a post-lesson interview, were very low. The middle- and high-school educators gave the lesson largely positive reviews. The elementary educators had strongly negative reactions to what they regarded as the insultingly low level of expectations for students and what they perceived as the teacher's condescending attitude toward his students.
- <sup>5</sup> Because No Child Left Behind requires schools to get all students to the same level of proficiency, schools with students who entered with high achievement levels have to make little or no improvement to meet NCLB targets. So staff in high-achievement schools may or may not know much about the process of improvement.

You can live in Thailand or Texas  
and take classes at

TEACHERS COLLEGE  
COLUMBIA UNIVERSITY

Enroll in online classes in areas  
such as educational technology,  
social studies, health, cognitive  
psychology and staff development.

Earn a certificate or credits that  
may be applied to a degree.

For information visit our website at  
<http://dlp.tc.columbia.edu> or call 888.633.6933

DISTANCE  
LEARNING  
PROJECT



TEACHERS COLLEGE  
COLUMBIA UNIVERSITY



## Display Freedom's Spread in Your Classroom

To learn more about Freedom in the  
World or about the work of Freedom  
House, go to [www.freedomhouse.org/  
research/index.htm](http://www.freedomhouse.org/research/index.htm).

To order a free, poster-sized Map of  
Freedom for your classroom, go to  
[www.freedomhouse.org/research/  
map.htm](http://www.freedomhouse.org/research/<br/>map.htm).



## AYP

(Continued from page 39)

groups that aren't achieving as well as they should. The requirements also seem to be spurring schools to take concrete actions to improve student achievement (CEP, 2005).

The complexities of AYP have important implications for schools and districts. First, many parents and other citizens do not realize the various reasons why schools could miss AYP targets and may unfairly assume that something is seriously wrong with their local school, even when that's not the case. Clearly, the AYP process will identify a lot of *low-achieving* schools, but just because a school has been identified doesn't make it *low performing*. Second, if too many schools fail to make AYP, it can undermine the credibility of the federal requirements and create incentives for states to lower cut scores or dumb down their tests. Third, if too many schools are cited as needing improvement, states and districts will not have enough resources, staff, time, and expertise to help the schools with the most serious educational problems.

The intent of No Child Left Behind is to greatly increase learning for all groups of students by highlighting schools and subgroups with legitimate educational needs and marshalling resources to boost their achievement. Underperforming schools need technical and financial assistance to carry out proven strategies for improving student learning. Over the long term, the President and the Congress should consider revising the adequate yearly progress requirements

of NCLB based on evidence from CEP's study, other research, and the experiences of states and school districts. □

## References

- American Federation of Teachers (n.d.). Use of performance indexes. Unpublished paper.
- Center on Education Policy (2005). *From the capital to the classroom: Year 3 of the No Child Left Behind Act*. Washington, D.C.: CEP.
- Florida Department of Education (August 17, 2004). "School grades, adequate yearly progress, and school recognition." Commissioner's presentation to the State Board of Education. Retrieved from [www.fldoe.org/meetings/2004\\_08\\_16/SchoolGrades\\_Pres.pdf](http://www.fldoe.org/meetings/2004_08_16/SchoolGrades_Pres.pdf).
- Frahm, R. (May 28, 2004). "Study predicts school failures." *Hartford Courant*.
- Helderman, R. S. and Mui, Y. Q. (September 25, 2003). View on schools' progress unclear. *Washington Post*.
- New Mexico State Department of Education (n.d.). *Consolidated state application accountability workbook*. Retrieved from [www.ed.gov/admins/lead/account/stateplans03/nmcsa.pdf](http://www.ed.gov/admins/lead/account/stateplans03/nmcsa.pdf).
- New York State Department of Education (February 2005). "Grade 4 English Language Arts (ELA) Test." Retrieved from [www.emsc.nysed.gov/osa/concht/05/gr4ela.htm](http://www.emsc.nysed.gov/osa/concht/05/gr4ela.htm).
- Olson, L. (December 8, 2004). "Taking root." *Education Week*.
- Oregon Department of Education (n.d.). "Key points about adequate yearly progress (AYP)." Retrieved from [www.ode.state.or.us/initiatives/nclb/docs/aypkeypoints.rtf](http://www.ode.state.or.us/initiatives/nclb/docs/aypkeypoints.rtf).
- Tennessee Department of Education (n.d.). "Proficiency ranges for TCAP achievement 2004-2005 CRT scores." Retrieved from [www.state.tn.us/education/assessment/tsach03nproflvl.php](http://www.state.tn.us/education/assessment/tsach03nproflvl.php).
- Virginia Board of Education (n.d.). *Consolidated state application: Amended accountability workbook*. Retrieved from [www.ed.gov/admins/lead/account/stateplans03/vacsa.pdf](http://www.ed.gov/admins/lead/account/stateplans03/vacsa.pdf).
- Wisconsin Department of Public Instruction. (September 17, 2004). "Preliminary AYP notification sent to schools and districts." News release.

### EDUCATORS B&B NETWORK

Educators Hosting Educators Since 1986

**\$34**

**A NIGHT  
FOR TWO!**

6,000 Locations Worldwide

A Few Comments from our members:

▶▶ **"Wonderful Hosts!"**

Ron & Susan, Seattle, WA #12105

▶▶ **"What a Delight!"**

Beth & Carl, Ann Arbor, MI #5898

▶▶ **"Amazing Hospitality!"**

Mary & Bud, Sacramento, CA #6785

**Visit our website today  
and start traveling!**

Box 5279 - Eugene OR 97405  
800-377-3480

[www.educatorstravel.com](http://www.educatorstravel.com)

**W**hen natural disasters strike at home or around the world, the AFT family has always responded with generosity. We saw that commitment again following the devastating tsunami that struck in December 2004.

With the help of AFT members, we are working together with our fellow trade unionists to provide the resources needed by educators, public employees, and healthcare workers in the countries affected by the tsunami. Education

### *Tsunami Relief... The need continues...*

International, the international organization of teacher unions, estimates that more than 10,000 teachers died or were greatly affected by the disaster.

Your tax-deductible contribution to the AFT Educational Foundation will be put to work directly by the AFT, Education International, Public Services International and the AFL-CIO's Solidarity Center for Humanitarian Assistance, where it is needed most. Contributions should be made payable to the AFTEF and sent to the AFT Educational Foundation, 555 New Jersey Avenue, N.W., Fourth Floor, Washington, DC 20001.

Please detach coupon and send it along with your tax-deductible contribution to AFTEF. Please note "Tsunami Relief" on the memo portion of your check.

Name \_\_\_\_\_  
Address \_\_\_\_\_  
City \_\_\_\_\_ State \_\_\_\_\_ Zip \_\_\_\_\_  
E-mail address \_\_\_\_\_  
Local \_\_\_\_\_







# Teaching *all* Reaching *all*



**TEACHING** is a profession that requires all educators to be current. QuEST offers 100 workshops conducted by national experts and excellent teachers from across the country.

**ALL** students deserve great teachers. QuEST gives educators a chance to engage in dialogue with national education experts and colleagues who grapple with the same issues they confront every day.

**REACHING** all students requires teachers to know about the latest research and best practices. Plenary sessions, seminars, exhibits and a resource and technology center will bring you up to date.

**ALL** teachers deserve quality professional development. Please join over 2800 of your colleagues from across the nation for:

**QuEST 2005**  
**JULY 7-10**  
**HILTON WASHINGTON**  
**WASHINGTON, D.C.**

FOR MORE INFORMATION, OR TO REGISTER ONLINE (MEMBERS ONLY), GO TO [www.aft.org/QuEST2005](http://www.aft.org/QuEST2005)



*A Union of Professionals*

**QuEST 2005**

Quality Educational Standards in Teaching