

**EDUCATION WEEK®**

# **Room to Maneuver**

**A Progress Report on the No Child Left Behind Act**

**December 14, 2005**

**edweek.org**

# TABLE OF CONTENTS

## **Dec. 14, 2005**

### **Room to Maneuver**

The federal government has responded to criticism of the No Child Left Behind Act by giving states greater flexibility in how to comply. Analysts worry that the law's purpose is being lost along the way.

### **Table: School Ratings and Student Performance**

States have a mixed record on getting their schools to meet student-achievement goals under the law.

### **Political Shifts Cloud Outlook for Renewal of Federal Education Law**

A changing cast of characters clouds the outlook for the scheduled reauthorization of the statute.

### **Actual Measure of 'Highly Qualified' Teachers Just Beginning to Come to Light Across Nation**

States report progress on meeting the law's mandates for having "highly qualified" teachers.

### **Survey Results: Where Are the 'Highly Qualified' Teachers?**

Individual states report data on the percent of classes taught by highly qualified teachers.

### **States Confront Definition of 'Proficient' as They Set the Bar for Lots of New Tests**

States face complexities in making sure they have similar test-scoring standards in every grade.

## **Nov. 30, 2005**

### **State Test Programs Mushroom as NCLB Mandate Kicks In**

You could describe it as the year of the tests. Twenty-three states are expanding their testing programs to additional grades this school year to comply with the federal No Child Left Behind Act.

### **Federal Review Puts State Tests Under Scrutiny**

By the end of November, 26 states will have undergone a "peer review" to determine whether their standards and tests meet the requirements of the federal No Child Left Behind Act.

### **Benchmark Assessments Offer Regular Checkups on Student Achievement**

School districts worried about how students will perform on end-of-the-year state tests are increasingly administering "benchmark assessments" throughout the year to measure students' progress and provide teachers with data about how to adjust instruction.

### **Not All Teachers Keen on Periodic Tests**

Across the country, school districts are adopting benchmark assessments to help teachers modify instruction over the course of a school year. Yet many teachers remain wary.

# Room to Maneuver

*The federal government has responded to criticism of the No Child Left Behind Act by giving states greater flexibility in how to comply. Analysts worry that the law's purpose is being lost along the way.*

By Lynn Olson

Since U.S. Secretary of Education Margaret Spellings announced last April that she would take a "common sense" approach to carrying out the No Child Left Behind Act, educators have been watching closely to see whether she keeps that promise.

The answer, according to observers, is yes—and no.

"It's definitely the right approach," said Scott R. Palmer, who helps lead the education policy team at the Washington-based law firm of Holland & Knight. "I think it remains a work in progress."

As the ambitious federal law nears its fourth anniversary on Jan. 8, people following its progress say Ms. Spellings' willingness to work with states has been a welcome shift from the relatively hard-nosed stance of her predecessor, Rod Paige. And it's helped defuse at least some criticisms of the law at the state and local levels.

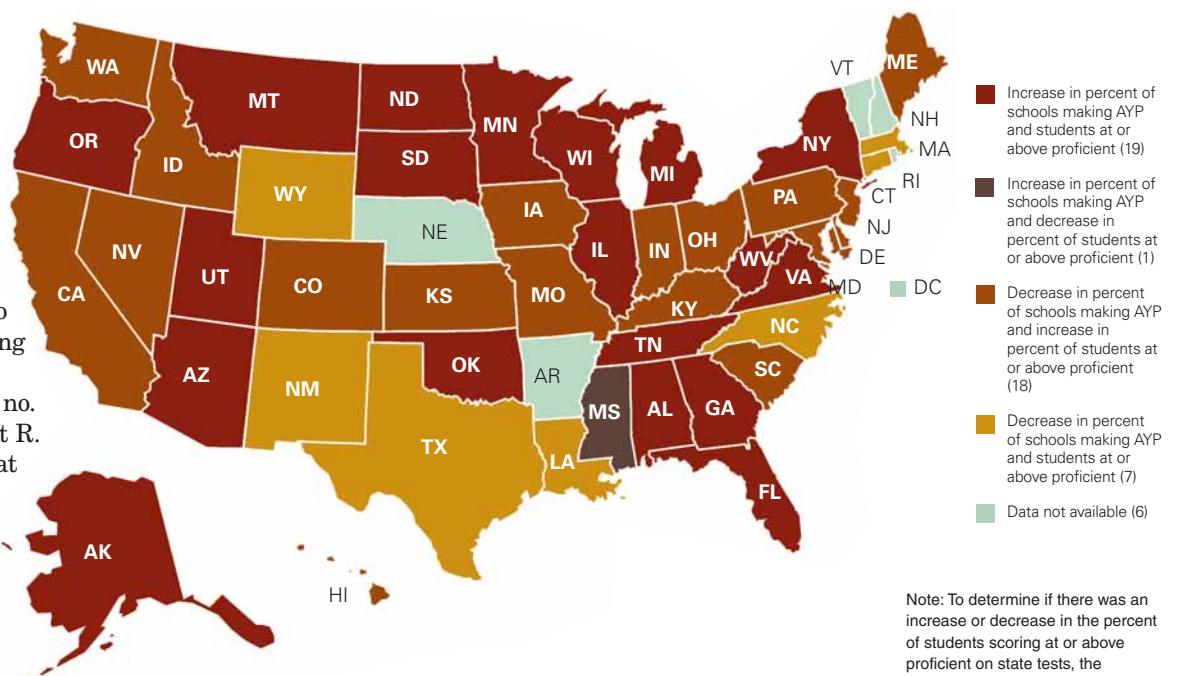
"You have to give a little credit where credit is due," said Reg Weaver, the president of the 2.7 million-member National Education Association, which has sued the secretary over the statute. "I'm not going to say what the secretary has done is not a step in the right direction, because I think it is. But still more needs to be done."

An *Education Week* analysis of amendments to state accountability workbooks, which detail how states are complying with the law's testing-and-accountability provisions, shows that much of the flexibility granted by the federal Department of Education since April has consisted of extending agreements already reached with some states to others.

But the secretary also has tried to respond to critics and fend off changes to the law itself by relaxing rules for measuring the progress of special education students, extending the deadlines for some states to meet the law's "highly qualified" teacher goals, and launching pilot programs that modify the law's school choice and tutoring requirements for schools in need of improvement.

Most recently, on Nov. 18, Ms. Spellings announced that she would permit up to 10 states to take part in a pilot program enabling them to calculate progress under the federal law based on "growth models" that track the gains of individual students over time.

"She's given the states better tools, or more opportunities, to do some things that would reduce the numbers of schools in need of improvement," said Dianne M. Piché, the executive director of the Citizens' Commission on Civil Rights, a Washington-based watchdog group that



## Inconsistent Data

In a comparison of data from the 2003-04 and 2004-05 school years, states that improved student performance on state-required exams did not always see a subsequent improvement in the percent of public schools making adequate yearly progress. States must raise proficiency targets at least every three years, under the NCLB law. So schools must continually increase the percent of students scoring "proficient" or higher in order to maintain a favorable rating.

Note: To determine if there was an increase or decrease in the percent of students scoring at or above proficient on state tests, the percentage-point change between 2004 and 2005 was averaged across 4th and 8th grade reading and mathematics. Some states average additional grades in order to determine AYP status. If a state did not give tests at grade 4 or 8, the EPE Research Center accepted test results from the next closest grade level. In Ohio, only 4th grade reading and mathematics were averaged.

SOURCE: Editorial Projects in Education Research Center, 2005

has been a strong supporter of the law. "To the extent that translates into a perception that NCLB is not so onerous and not so punitive, I think it does help in the field, and it does help with the public perception."

But in the long run, Ms. Piché said, the changes won't eliminate the deeper problem of a "large, gaping achievement gap" based on race and poverty, which the law was meant to address.

Her group has raised concerns, for example, that the growth-models program could lower expectations for the most disadvantaged youngsters.

## Leeway Extended

Championed by President Bush and approved overwhelmingly by Congress in late 2001, the law is a reauthorization of the Elementary and Secondary Education Act, the landmark 1965 measure that greatly expanded federal aid for K-12 education.

But the No Child Left Behind law goes beyond previ-

CONTINUED ON PAGE S2

## School Ratings and Student Performance

	AYP Status		In-Need-of-Improvement Status		Percentage of students who scored at or above proficient on state tests in 2005							
	Percent of schools that made AYP 2005	Percentage-point change from 2004 <sup>3</sup>	Percent of schools identified as in need of improvement 2005 <sup>4</sup>	Percentage-point change from 2004 <sup>3</sup>	4th grade reading	Percentage-point change from 2004	8th grade reading	Percentage-point change from 2004	4th grade math	Percentage-point change from 2004	8th grade math	Percentage-point change from 2004
Alabama	53%	+22	34%	+28	83%	+6	69%	+11	73%	+1	63% <sup>5</sup>	+38 <sup>5</sup>
Alaska	59	+1	38	+2	79 <sup>5,6</sup>	+5 <sup>5,6</sup>	80 <sup>5</sup>	+13 <sup>5</sup>	76 <sup>5,6</sup>	+3 <sup>5,6</sup>	62 <sup>5</sup>	-2 <sup>5</sup>
Arizona	86	+3	9	-1	63 <sup>6,7</sup>	+1 <sup>6,7</sup>	63 <sup>7</sup>	+19 <sup>7</sup>	71 <sup>6,7</sup>	+11 <sup>6,7</sup>	59 <sup>7</sup>	+36 <sup>7</sup>
Arkansas	—	—	24	-6	52 <sup>7</sup>	-17 <sup>7</sup>	57 <sup>7</sup>	+5 <sup>7</sup>	50 <sup>7</sup>	-15 <sup>7</sup>	33 <sup>7</sup>	+1 <sup>7</sup>
California	62	-4	19	+1	47	+8	39	+6	50	+5	37 <sup>6</sup>	+4 <sup>6</sup>
Colorado	75	-4	6	+1	86	-1	86	0	90	+3	75	+5
Connecticut	80	-1	16	+2	67	-2	75	-2	79	-1	76	-1
Delaware	74	-2	21	-4	84 <sup>6</sup>	+2 <sup>6</sup>	78	+8	79 <sup>6</sup>	+1 <sup>6</sup>	53	+3
District of Columbia	45	+8	35	-6	—	—	—	—	—	—	—	—
Florida	36	+12	32	+1	71	+1	44	-1	64	0	59	+3
Georgia	82	+2	17	-3	87	+8	83	-2	75	-1	69	-4
Hawaii	34	-18	48	-1	52 <sup>6</sup>	+5 <sup>6</sup>	38	-1	29 <sup>6</sup>	+1 <sup>6</sup>	21	0
Idaho	57	-25	15	+3	87	+4	82	0	90	+7	69	+3
Illinois	73	+1	17	0	67 <sup>6</sup>	+2 <sup>6</sup>	73	+6	79 <sup>6</sup>	0 <sup>6</sup>	54	0
Indiana	60	-16	5	0	75 <sup>6</sup>	+1 <sup>6</sup>	67	+2	73 <sup>6</sup>	+2 <sup>6</sup>	71	0
Iowa	91	0	6	+2	79 <sup>6</sup>	+3 <sup>8</sup>	72 <sup>8</sup>	+3 <sup>8</sup>	81 <sup>8</sup>	+4 <sup>8</sup>	75 <sup>8</sup>	+3 <sup>8</sup>
Kansas	91	-2	1	0	78 <sup>6</sup>	+6 <sup>6</sup>	77	+2	85	+5	68 <sup>6</sup>	+3 <sup>6</sup>
Kentucky	74	-2	11	0	68	+1	62 <sup>5</sup>	+2 <sup>5</sup>	45 <sup>5</sup>	-3 <sup>6</sup>	36	+3
Louisiana	84	-8	13	+7	64	+1	50	0	61	+4	51	-9
Maine	77	-3	8	0	53	+3	44	+7	39	+7	29	+7
Maryland	75	-5	18	-1	81	+6	66	+3	76	+7	52	+6
Massachusetts	57	-16	24	+3	50	-6	66 <sup>6</sup>	-2 <sup>6</sup>	40	-2	39	0
Michigan	88	+10	13	-1	82	+2	73 <sup>6</sup>	+12 <sup>6</sup>	73	0	62	-1
Minnesota	87	+12	4	+2	78 <sup>6</sup>	+5 <sup>6</sup>	74 <sup>6</sup>	+4 <sup>6</sup>	78 <sup>6</sup>	+7 <sup>6</sup>	76 <sup>6</sup>	+9 <sup>6</sup>
Mississippi	89	+12	9	+1	89	+1	57	-5	79	-1	53	-7
Missouri	65	-12	7	0	35 <sup>6</sup>	0 <sup>6</sup>	33 <sup>6</sup>	+1 <sup>6</sup>	43	+3	16	+2
Montana	93	+7	9	+1	75	+10	64	+6	57	+12	63	-1
Nebraska	—	—	—	—	85	—	86	—	88	+3	82	+5
Nevada	47	-16	29	+7	45 <sup>6</sup>	0 <sup>6</sup>	51	+1	51 <sup>6</sup>	+6 <sup>6</sup>	49	+1
New Hampshire	53 <sup>2</sup>	— <sup>2</sup>	42 <sup>2</sup>	— <sup>2</sup>	— <sup>2</sup>	— <sup>2</sup>	— <sup>2</sup>	— <sup>2</sup>	— <sup>2</sup>	— <sup>2</sup>	— <sup>2</sup>	— <sup>2</sup>
New Jersey	61	-11	25	+1	82	-2	72	-1	80	+6	62	-2
New Mexico	47	-19	30	+13	52 <sup>5</sup>	+3 <sup>5</sup>	51 <sup>5</sup>	-5 <sup>5</sup>	39 <sup>5</sup>	-16 <sup>5</sup>	24 <sup>5</sup>	-24 <sup>5</sup>
New York	80	+5	18	0	70	+7	48	+1	85	+6	56	-3
North Carolina	58	-13	9	+2	82	0	88	0	92	-2	84	0
North Dakota	91	+1	4	-1	75	-6	72	0	79	+14	65	+19
Ohio	76	-7	13	+1	77 <sup>5</sup>	+6 <sup>5</sup>	79	—	66	0	60	—
Oklahoma	97	+22	7	-1	79 <sup>6</sup>	+3 <sup>6</sup>	81	-1	84 <sup>6</sup>	+6 <sup>6</sup>	76	-1
Oregon	68	0	26	+6	86 <sup>6</sup>	+4 <sup>6</sup>	63	+4	86 <sup>6</sup>	+5 <sup>6</sup>	64	+5
Pennsylvania	81	-6	13	-2	64 <sup>6</sup>	+1 <sup>6</sup>	64	-5	69 <sup>6</sup>	+8 <sup>6</sup>	63	+5
Rhode Island	44 <sup>2</sup>	— <sup>2</sup>	18 <sup>2</sup>	— <sup>2</sup>	— <sup>2</sup>	— <sup>2</sup>	— <sup>2</sup>	— <sup>2</sup>	— <sup>2</sup>	— <sup>2</sup>	— <sup>2</sup>	— <sup>2</sup>
South Carolina	47	-9	15	-5	36	-2	30	+3	41	+5	23	+1
South Dakota	84	+5	15	0	87	0	79	+1	82	+4	69	+3
Tennessee	93	+7	9	0	87	+7	87	+6	87	+7	87	+4
Texas	87	-7	3	+1	79	-6	83	-6	81	-5	61	-5
Utah	87	+3	1	0	78	+2	77	-1	75	+2	73	+3
Vermont	86 <sup>2</sup>	— <sup>2</sup>	21 <sup>2</sup>	— <sup>2</sup>	— <sup>2</sup>	— <sup>2</sup>	— <sup>2</sup>	— <sup>2</sup>	— <sup>2</sup>	— <sup>2</sup>	— <sup>2</sup>	— <sup>2</sup>
Virginia	81	+6	6	0	77 <sup>6</sup>	+6 <sup>6</sup>	76	+4	88 <sup>6</sup>	+1 <sup>6</sup>	81	+1
Washington	80	-4	9	0	80	+5	69 <sup>6</sup>	+9 <sup>6</sup>	61	+1	51 <sup>6</sup>	+4 <sup>6</sup>
West Virginia	83	+12	5	0	81	+8	80	0	75	+6	70	+1
Wisconsin	98	+2	2	0	81	0	84	+5	71	-2	72	+7
Wyoming	80	-12	4	-1	47	0	39	-2	39	0	38	-2
<b>Total<sup>1</sup></b>	<b>74</b>	<b>-1</b>	<b>14</b>	<b>+1</b>	<b>—</b>	<b>—</b>	<b>—</b>	<b>—</b>	<b>—</b>	<b>—</b>	<b>—</b>	<b>—</b>

ous reauthorizations by holding all public schools accountable for raising all students to academic proficiency by 2013-14. It requires states to test students in reading and mathematics annually in grades 3-8 and at least once during high school, beginning this school year.

To make adequate yearly progress, or AYP, which is a key measurement under the law, schools and districts must meet annual performance targets, based primarily on those test scores. The targets apply to their student populations as a whole and to subgroups of students who are poor, speak limited English, have disabilities,

or are members of racial or ethnic minorities.

Schools and districts receiving federal Title I money for disadvantaged students are subject to increasingly serious sanctions if they miss the targets for two or more years. Initially, they must permit students to transfer to a higher-performing public school; then, they must offer eligible students free tutoring from a public or private provider. Eventually, such schools and districts are subject to "corrective action," including, potentially, school closure.

The provisions on testing and accountability, in particular, have stirred resistance from states such

as Connecticut and Utah, which have complained that federal funding is inadequate and that the federal government has overstepped its bounds by dictating the details of state accountability policies. Like the NEA, Connecticut is pursuing a court challenge over the law. Utah passed legislation this year that says its own state accountability system takes precedence over the federal law.

Since last spring, Secretary Spellings—who took office in January after serving as a White House domestic-policy aide during Mr. Bush's first term—has granted new leeway in five specific areas by:

# We love testing!

Notes: A dash (—) indicates that data were not available or not provided to the EPE Research Center prior to deadline, or, in U.S. row, that a total was not appropriate. Some 2005 percentages are based on preliminary data. Some 2004 student-achievement data were gathered from [www.schoolmatters.com](http://www.schoolmatters.com).

<sup>1</sup>Total is based on states where data were available.

<sup>2</sup>New Hampshire, Rhode Island, and Vermont suspended much of their testing programs during the 2004-05 school year as they transitioned to a fall testing cycle. For 2005, the percentage of elementary and middle schools making AYP is primarily based on attendance. High school results for 2005 are based on assessment data and graduation rates. For these states, the table only reflects the percentage of high schools making adequate yearly progress and identified as in need of improvement, so data are not comparable to those of prior years.

<sup>3</sup>Negative values in these columns indicate a decline in the percent of schools making AYP or identified as in need of improvement.

<sup>4</sup>All states must include Title I schools in their designation of schools "in need of improvement." Federal law allows states to choose whether non-Title I schools are assigned a school improvement status and whether federal consequences apply to those schools. As a result, some schools rated for AYP may not receive a school improvement designation. Percentage was calculated by dividing the total number of schools identified as in need of improvement by the total number of schools rated for AYP.

<sup>5</sup>State implemented a new assessment in 2005; results prior to 2005 may not be comparable.

<sup>6</sup>State did not offer tests at grades 4 or 8. The EPE Research Center accepted test results from the next closest grade level.

<sup>7</sup>In Arizona and Arkansas, 2005 results represent the beginning of a new trend line; results prior to 2005 may not be comparable.

<sup>8</sup>Iowa 2004 student-achievement results represent the average between scores from the 2002-03 and 2003-04 school years. The 2005 results are based on the 2004-05 school year only.

SOURCE: Editorial Projects in Education Research Center, 2005

- Permitting states to calculate the progress of some students with disabilities using alternative assessments that are not pegged to a grade-level standard, the so-called "2 percent rule";

- Allowing some schools and districts heavily affected by Hurricanes Katrina and Rita to create separate subgroups of displaced students for this school year only, and not count those students in any other subgroup in calculating progress under the federal law;

- Approving a pilot program in Virginia that lets several districts reverse the order in which public school choice and tutoring are offered to students in schools identified for improvement. The Education Department also has permitted a handful of urban districts to continue providing their own tutoring even if the districts themselves have been identified for improvement;

- Giving states an extra year to ensure that all of their classes in the core academic subjects are taught by "highly qualified" teachers, as long as they can show a "good-faith effort" in meeting the law's requirements for such teachers; and

- Announcing the pilot program permitting as many as 10 states to use a growth model to measure school and district progress under the law.

But what's most striking is the number of states that have taken advantage of already existing flexibility under the law to make it easier for schools and districts to make adequate progress.

## Design Matters

For example, 23 states now require at least 40 students to be in a subgroup before it counts for accountability purposes. At least eight states also require each subgroup to constitute a minimum percent of the overall student enrollment before it counts in calculating AYP, until the subgroup reaches a certain size. And eight states have larger subgroup requirements for students with disabilities and those learning English.

The vast majority of states also now apply a statistical test known as a "confidence interval" that makes it easier for schools and districts to make AYP by creating a margin of error around performance results.

Half the states have created timelines for getting all students to the proficient level by 2013-14 that postpone steep increases in the improvement targets until closer to that deadline year set by the law.

Such design decisions can lead to "pretty big differences" in the percentages of schools that make adequate progress, according to Andrew C. Porter, the director of the Learning Sciences Institute at Vanderbilt University

CONTINUED ON PAGE S4



**"We now have reliable and valid data to differentiate instruction and are able to engage in using innovative structures to support students with similar needs. In the district, we are refining our curriculum to better achieve our goals, based on insights we have from the testing results."**

— Ginger Hopkins, Assistant Superintendent, Beaufort County, South Carolina

**"NWEA's MAP test provides 'just-in-time' data that we cannot get in other places. Previously, results came back too late to make many changes in our instructional program."**

— Debra Hill, Superintendent, West Northfield, Illinois

**"When new students arrive, I can assess their skills and design appropriate IEP goals. We feel better prepared to teach students according to their academic needs."**

— Sheri Onion, Special Education Teacher, East Noble SC, Indiana

Specifically, we love NWEA testing. It gives us timely, highly accurate achievement data aligned to state standards. We know what's working and what's not from an individual student to an entire district.

Our kids love it too. It's adaptive, so they don't take tests that are too easy or too hard. We see precise results for all students. And we get specific concepts to guide instruction, curriculum and school improvement. NWEA inspires administrators, principals, teachers, parents and students alike.

NWEA is a non-profit organization dedicated to helping every child learn. That's why we're members. See what it's already meant for more than two million students at [www.NWEA.org/DataTalk](http://www.NWEA.org/DataTalk).

  
Northwest Evaluation Association  
Partnering to help all kids learn

CONTINUED FROM PAGE S3

in Nashville, Tenn. A simulation he and others conducted, based on 2004 test data from Kentucky, found that the proportion of schools there making AYP varied from 59 percent to 89 percent, based on the design decisions in those three areas: subgroup size, the use of a confidence interval, and the timeline trajectory for getting all students to proficiency.

At least 31 states also chose to identify districts for improvement under the federal law only if they miss their targets for two years in a row across all of the grade spans tested: elementary, middle, and high school, or elementary/middle and high school.

In a report released last month on 2005 amendments to state accountability plans, prepared for the Washington-based Council of Chief State School Officers, authors William J. Erpenbach and Ellen Forte concluded that, in the long run, “the accumulation of amendments and ‘flexibility’ may result in educational accountability systems that lack any real connection to the achievement goals they were supposed to realize.”

Jack Jennings, the president of the Center on Education Policy, a Washington think tank, indicated that he sees the extra leeway that Ms. Spellings granted this year as less significant than the use states have made of previously existing flexibility under the NCLB law.

“What is troubling is that what has been

granted—especially with confidence intervals and with the size of subgroups—raises questions about whether the implementation of the law has become so loose now that the law may be losing some of its meaning,” Mr. Jennings said.

Last month, the Education Department released a report designed to show the “variety of fair and reliable methods” states are using to comply with the law, and why the department decided to accept such diverse approaches.

“Because each state is unique, no two state accountability plans are identical,” the report said. “While approved changes to state accountability plans are not uniform across the states, our criteria for evaluating and approving such changes are uniform. They reflect the department’s commitment to maximizing accountability while minimizing error in measuring school performance.”

Meanwhile, the extra leeway “has reduced the political temperature to a degree,” added Mr. Jennings, a former staff director of the U.S. House of Representatives’ education committee under a Democratic majority.

“But there’s still strong opposition to the law, and we’re going to see that in January and February when state legislatures reconvene,” he added.

Some of the most contentious issues—such as how to measure progress under the law—likely will be taken up during the law’s reauthorization, scheduled for 2007. (See related story, this page.)

Of the 45 states and the District of Columbia with data available, the percentage of schools

that made AYP in 2005 rose in 21 states, fell in 25 compared with 2004, according to an analysis conducted by the Editorial Projects in Education Research Center for *Education Week*. The percentage of schools in need of improvement declined in 19 states, compared with 2004. (See chart, Page S2.)

While the proportion of students scoring at or above the proficient level on state tests also rose in many states this past year, the mixed national picture on AYP may have as much to do with how each state calculates progress as on overall test-score trends.

### Bar ‘Keeps Going Up’

Despite the leeway granted to states, many experts predict that it will be harder for schools to make their performance targets in the future. That’s because starting this school year, nearly every state will conduct testing in reading and math at all the required grade levels.

That change makes it more likely schools will meet the minimum subgroup requirements for different groups of students. In addition, many states in 2004-05 raised for the first time their targets for the percent of students who must score at the proficient level, and those targets will continue to rise.

“As you look at the content of the requests that have been granted, will that be enough to take the pressure off as you look at next year’s

## Political Shifts Cloud Outlook for Renewal of Federal Education Law

By Michelle R. Davis

Washington

A shifting political landscape could significantly affect the upcoming congressional review of the No Child Left Behind Act, even as states and school districts continue to grapple with the requirements of the nearly 4-year-old federal law.

The law is slated for reauthorization in 2007, and that process is starting now, at least informally, as education groups poll their members for ideas, and policymakers consider proposals for change that have already been put forward.

But all sides are peering into a murky future. The players and alliances surrounding the law’s renewal are sure to differ in some ways from the lineup in 2001, when a bipartisan group of legislators hammered out the current provisions with the Bush administration.

A wild card is the political stature of President Bush in the remaining years of his presidency. The No Child Left Behind law, which drew on Mr. Bush’s approach to school improvement while he was governor of Texas, was a centerpiece of his first-term agenda. It passed Congress at a time of national harmony after the terrorist attacks of Sept. 11, 2001, and Mr. Bush signed the measure in January 2002 with the support of prominent Democrats.

Such political togetherness now is rare, the president’s popularity has fallen, and pundits have begun tagging

Mr. Bush with the lame-duck label.

Such factors could cause the process to bog down before a new version of the law—itsself a renewal of the Elementary and Secondary Education Act, first passed in 1965—is finally approved. After all, the last reauthorization had been due to occur under President Bill Clinton in 1999.

“It might start now, but it’s going to take years and years and years,” Vic Klatt, a lobbyist for the Washington-based firm Van Scoyoc Associates and a former aide to Republicans on the House education committee, said of the NCLB reauthorization. “To make big changes, ... all the stars have to be aligned.”

### Changes in the Lineup

Ever since its enactment, the law’s supporters have fought off legislative challenges to the statute itself. Instead, the U.S. Department of Education has promoted the use of regulatory flexibility allowed under the law to help deal with problems in its implementation.

But as reauthorization hearings get under way—some likely as soon as 2006—new personalities may alter the terrain. Those changes may extend all the way to the White House, depending on how long the process takes.

During the last go-round, the “big four”—then-Chairman Edward M. Kennedy, D-Mass., of the Senate education committee and the ranking commit-

tee Republican, Sen. Judd Gregg, R-N.H., as well as House education committee Chairman John A. Boehner, R-Ohio, and ranking committee Democrat George Miller of California—worked together with the White House to craft the bill and push it through Congress.

But Rep. Boehner’s term as chairman expires at the end of 2006, and a new committee leader—possibly Rep. Howard “Buck” P. McKeon, R-Calif., if the Republicans keep control of the House—will take his spot.

The Senate education committee chairmanship has already changed hands twice since 2001, passing from Sen. Kennedy to Sen. Gregg with a change of party control in the Senate, and more recently to Sen. Michael B. Enzi, R-Wyo. Mr. Enzi, a strong supporter of public schools who has particular concerns about the No Child Left Behind law’s effect on rural areas, could lead the reauthorization in new directions.

Sen. Enzi said last month that it was “too early to tell yet” how reauthorization would go. “We have 39 reauthorizations to come up before that,” he said.

A spokeswoman for Mr. Boehner stressed that the process was ongoing as lawmakers evaluate how the current law is being implemented and look to the future.

“The reauthorization process entails a lot more than writing a bill and passing it through committee,” said Alexa Marro, a spokeswoman for Republicans on

the House Education and the Workforce Committee.

New chairmen in both the House and the Senate may vary in their commitment to the law’s original tenets, analysts say.

The personnel shifts “help increase the likelihood of changes to the law, because there’s less pride of authorship,” said Joel Packer, a lobbyist for the 2.7 million-member National Education Association, which sued the Department of Education last April over the law.

### Seeking Input

The two key Democratic players, Sen. Kennedy and Rep. Miller, have held firm on the core mission of the law—despite pressure from some liberal-leaning groups and misgivings about what they see as low funding levels. Depending on the outcome of the 2006 midterm elections, which could shift party control of the House, the Senate, or both, Sen. Kennedy and Rep. Miller could end up at the helms of the education committees.

“Mr. Miller wants to hear from people all over the country—educators, parents, and experts, among others—about their experience with the law,” said Tom Kiley, a spokesman for Rep. Miller, who stressed that reauthorization was still a while off. “He is actively seeking their input as reauthorization nears.”

When the law was first enacted four years ago, many conservative Republicans had concerns about its far-reaching

AYP results?" said Michael A. Resnick, the associate executive director of the National School Boards Association, based in Alexandria, Va. "For many districts, the bar keeps going up."

Others complain about a lack of consistency and transparency on the flexibility some states have gotten compared with others.

David Griffith, the director of governmental affairs for the National Association of State Boards of Education, also based in Alexandria, said its members have asked NASBE to start compiling and collecting such information "so that there's more transparency and they can learn from each other."

Utah state Rep. Margaret Dayton, the Republican who led her state's opposition to the federal act's reach into schools, said the prospect of greater flexibility for states has not been realized in Utah's case. Federal officials have denied several waivers requested by the state, she said. From her perspective, she said, the awarding of NCLB waivers appears "arbitrary and capricious."

"There's so much frustration across the board, whether it's from traditionally conservative Utah or traditionally liberal Connecticut," Mrs. Dayton said. "You have plenty of room for frustration on this."

### Looking to 2007

With the law up for reauthorization in 2007, many education groups are gearing up to try to

change it. For example, a 61-group coalition—including education, religious, and civil rights organizations—has signed on to a statement calling for Congress and the Bush administration to make 14 specific changes to the act.

Other education groups, including the Council of Chief State School Officers and the National Association of Secondary School Principals, also have put proposals on the table or are planning to do so.

"My sense is that every group in the alphabet soup will have proposals that are cast as improvements," said Susan Traiman, the director of education and workforce policy for the Washington-based Business Roundtable, which has been a strong supporter of the law, "and some will be improvements, and some will not be in the spirit of the law."

Yet despite the many proposals to change the law, many educators acknowledge that it has provided leverage for focusing on the needs of poor and minority youngsters.

Brent Walker, the principal of the 300-student Haverhill Cooperative Middle School in North Haverhill, N.H., and a member of the NASSP's No Child Left Behind Task Force, said: "It's a way to have conversations with teachers and parents about tough things that needed to happen and jump-start change." ■

*Staff Writer Michelle R. Davis contributed to this report.*

*States are taking advantage of regulatory flexibility under the federal law to avoid having schools and districts fall short of standards for adequate progress.*

federal mandates, but were willing to give President Bush the benefit of the doubt. Now, some of them may no longer be willing to follow his lead.

Many lawmakers, meanwhile, have been bombarded in their home states by criticisms of the federal law's effects in such areas as state spending and the amount of testing in local schools.

"I don't think five people get to sit in a room and work it out this time," said Bruce Hunter, the chief lobbyist for the Arlington, Va.-based American Association of School Administrators, referring to the next version of the law.

### Proposals Vary

When it comes to specifics on how the law should be changed, many groups are pushing for the use of "growth models" to measure schools' progress.

Instead of focusing on the percent of students who score at or above an absolute level of performance, such models track the progress that individual students make from one year to the next.

On Nov. 18, U.S. Secretary of Education Margaret Spellings announced that up to 10 states could participate in a pilot project that uses growth models under the NCLB law, in part to help inform the reauthorization debate.

"The conversation about growth models is clearly going to be a very important part of the reauthorization discussion," said Ross Wiener, the policy director for the Education Trust, a Washington-based research and advocacy group that strongly supports the law.

Others predict that Congress will re-

visit, and possibly tighten, the law's teacher-quality provisions next time around. And, those observers say, the lawmakers will have to address continuing tensions about how to include students with disabilities and those learning English in state systems of testing and accountability.

President Bush, for his part, has sought to extend the No Child Left Behind Act's provisions further into high schools. So far, there's been little appetite on Capitol Hill to do so. But recent results from the National Assessment of Educational Progress, underscoring the poor reading performance of 8th graders, may amplify the drumbeat for improvements at the secondary school level.

In addition, many people are pushing for a new look at the law's consequences for schools and districts identified as needing improvement. Some want to target the sanctions—most notably, requirements for public school choice and for tutoring services—solely to the student subgroups that have failed to meet annual performance targets. Others want to reverse the order of the sanctions, so that schools can offer students free tutoring before giving them the option of transferring to higher-performing schools.

Other policy experts want to add incentives that would encourage districts to provide parents with more choices for students in low-performing schools. And many want to build on the law's requirement that schools in need of improvement write school improvement plans, to ensure that those plans actually have some teeth and that schools have the

support to carry them out.

"While it's important to acknowledge that schools need to improve, it's also important to have a good process in place to help them," Mr. Wiener said.

### Getting It Done

Well ahead of the scheduled reauthorization, some Republicans have introduced bills calling for changes to the law.

Among them is Rep. Lee Terry of Nebraska, who has introduced broad legislation that would loosen some of the NCLB requirements. Among other changes, the bill would allow states to use localized tests to assess students, and not have to report test results of some English-language learners to the federal government.

Rep. Terry belongs to the Republican Study Committee, a group of more than 100 of the current 230 GOP members of the House. The group is organized to promote an agenda that includes limiting the reach of the federal government. Given its numbers, it could play a pivotal role in the reauthorization process.

On the other side of the Capitol, Sen. George Allen, R-Va., who is said to be considering a run for his party's 2008 presidential nomination, has been a vocal critic of the law's demands. Earlier this year, he introduced a bill to give states—like Virginia—with strong accountability systems some relief from the law's requirements.

Two other Republicans, Sens. Susan M. Collins and Olympia J. Snowe, both of Maine, have introduced legislation to amend the No Child Left Behind Act by providing greater local control and a

number of flexibility measures. Those provisions would include new ways of meeting the mandate for "highly qualified" teachers and of rating the academic performance of students with disabilities.

Meanwhile, President Bush's political clock is ticking. By early 2009, his second term will be over, and if the reauthorization takes that long, another president will be calling the shots.

"A lot of folks in the administration would love to get it done on schedule because it really is one of the president's primary domestic legacies," said Eugene W. Hickok, a former deputy secretary at the Department of Education who left in January and is now a senior policy director at the Washington lobbying firm Dutko Worldwide. "Before I left, the folks on the Hill had the same sentiment, but it's awfully tough, and reauthorizations seldom happen on schedule."

Even if the reauthorization is completed on President Bush's watch, some analysts and lobbyists say it will be hard for him to hold on to the law's core mission.

And if it is pushed back beyond the 2008 elections, all bets are off, they say, given a new president's likely wish to put his or her stamp on the premier federal legislation in K-12 education.

"Whoever is the next president, they would likely be less supportive of No Child Left Behind," Mr. Packer of the NEA predicted. "It's in the president's advantage to get it done while he's still president." ■

*Senior Editor Lynn Olson contributed to this report.*

# Actual Measure of 'Highly Qualified' Teachers Just Beginning to Come to Light Across Nation

*First and second rounds of reported data based largely on guesswork*

By Bess Keller

If the No Child Left Behind law's prescription for "highly qualified" teachers had worked out the way it appears on paper, states would have gotten a good look at how far they had to go as far back as 2003. Then, as envisioned by Congress' 2001 overhaul of the Elementary and Secondary Education Act, they would have marched steadily toward the goal of outfitting each core-subject classroom with a teacher who meets the NCLB standard by the end of this school year.

Instead, many officials and observers say, the first round of data from states on the proportion of classes taught by highly qualified teachers was largely guesswork, and the second round was not much better. Only the statistics due Jan. 1 will tell a reliable story, officials in several states say.

"Every year, it's getting better," ventured Steven Olson, who is in charge of the teacher-qualification figures for the Rhode Island education department. "I'm very comfortable with it this year," he said of the data's reliability.

Keith Rheault, the state schools chief in Nevada, reports similar progress. "I can document all our [2004-05] numbers," he said, unlike those in the preceding two years.

Even with the improvement, the 2004-05 numbers are expected to remain inaccurate in at least some states. A report ordered by Congress and released last month found that "several limitations on the quality and precision of state-reported data make it difficult" to get good figures.

In fact, nine states don't even collect data on the percentage of core academic classes taught by highly qualified teachers, which is required by the U.S. Department of Education. Instead, they report the data by teachers or full-time equivalencies. States have had to submit highly qualified statistics to federal officials as part of their applications for federal education aid. The figures for the 2004-05 school year are not due until Jan. 1, but 22 states and the District of Columbia had provided them to the Editorial Projects in Education Research Center by mid-November. Eighteen other states provided 2003-04 data.

Of the 22 states and the District with 2004-05 data, 14 reported that at least nine out of 10 classes in their schools were taught by highly qualified teachers that year. The number rises to 25 if the states with 2003-04 data are included. Most states posted improvements compared with the statistics submitted to the U.S. Department of

Education three years ago.

In addition to mandating a highly qualified teacher in every classroom, the law calls on states to ensure that students in high-poverty schools have the same access to highly qualified teachers as do other students. Twelve of the 22 states and the District with 2004-05 data report that the percentage of classes in high-poverty schools taught by highly qualified teachers is better than or within 2 percentage points of the statewide figure. When states with 2003-04 data are included, 20 report the percentage of classes taught by highly qualified teachers in high-poverty schools is better than or close to the statewide average.

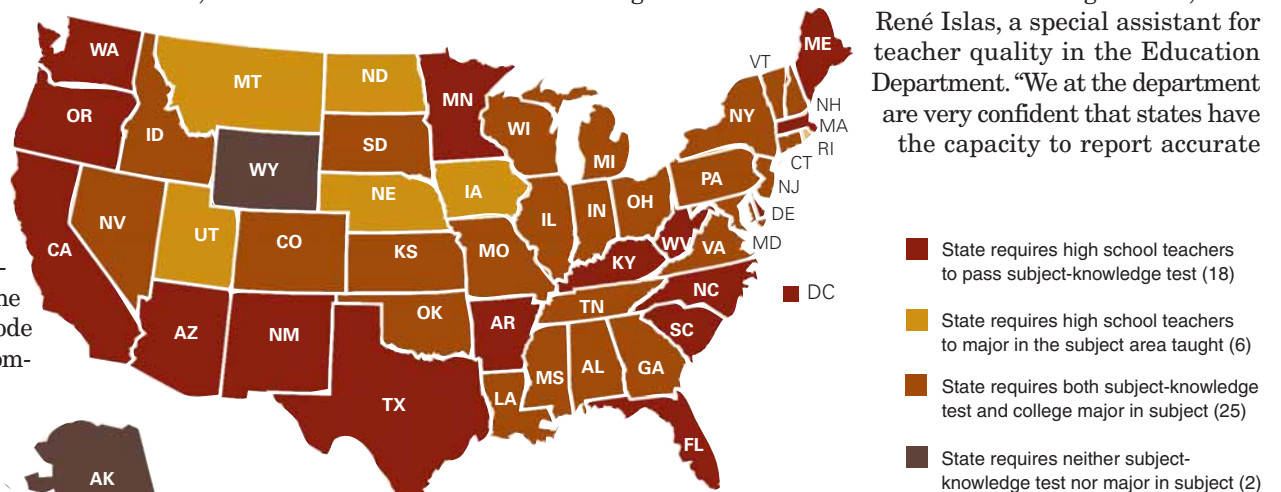
Given the state of data collection over the past three years, it is hard to know what the improved numbers represent, state officials say. On the one hand, districts and states have been making ef-

classroom skill—limits serious improvement even without implementation problems.

To meet the "highly qualified" standard, each teacher of a core subject must have a standard license from the state and demonstrate knowledge of the subject taught. New teachers have to do that by taking and passing tests in the subjects they teach or completing college majors in them. Teachers who were in the classroom three years ago, soon after the federal law was enacted, may go those routes or take an alternative one devised by their states within federal guidelines. Almost all the states now offer such alternatives, though they vary considerably in their requirements.

Federal officials say that their oversight of teacher quality has stepped up over the past year.

"We've monitored 35 states thus far, providing them with a lot of guidance," said René Islas, a special assistant for teacher quality in the Education Department. "We at the department are very confident that states have the capacity to report accurate



## Knowing What You Teach

According to the No Child Left Behind law, by the end of this school year, new teachers of core academic subjects must demonstrate knowledge of the subjects they teach by passing subject-knowledge tests or by completing subject-area majors. Almost every state now has one of those requirements in place.

forts to strengthen their teacher corps. On the other, what it takes for a veteran teacher to be deemed highly qualified has not been clear in some states until very recently. Perhaps the higher numbers show that more teachers who already had the background to be considered highly qualified have at last been counted.

## More Oversight

Many advocates of raising teacher quality say that federal officials have failed to push the law's teacher-quality agenda sufficiently and left states without the guidance or support they needed to do a better job.

Some also charge that what they see as the law's narrow definition of teacher quality—it emphasizes knowledge of subject matter over

data ... due at the beginning of the year."

In an Oct. 21 letter, U.S. Secretary of Education Margaret Spellings shifted away from a focus on the looming deadline and emphasized the need for progress. She said that states that want a year's reprieve from the threat of losing federal funds will have to show evidence they have been building the systems needed to take responsibility for the quality of their teaching forces. The deal also requires states to map out how they intend to move forward and subject their plans to the scrutiny of federal officials.

"If they haven't been attracting teachers to hard-to-staff schools," Mr. Islas said by way of example, "we'll want to see that they have made efforts to provide incentives to go there."

The new tack has heartened some critics of the

CONTINUED ON PAGE S8





*If they meet a set of criteria showing progress, states can get a year's reprieve in meeting the requirement for 'highly qualified' teachers.*

### Where Are the 'Highly Qualified' Teachers?

Percent of classes taught by "highly qualified" teachers (2004-05)<sup>1</sup>

	Statewide	High-poverty schools	Low-poverty schools
Alabama	77% <sup>2</sup>	68% <sup>2</sup>	79% <sup>2</sup>
Alaska	32	23	30
Arizona	96 <sup>2</sup>	96 <sup>2</sup>	96 <sup>2</sup>
Arkansas	—	—	—
California	74	65	81
Colorado	95	94	97
Connecticut	99 <sup>3</sup>	98 <sup>3</sup>	99 <sup>3</sup>
Delaware	73 <sup>3</sup>	—	—
District of Columbia	52	65	49
Florida	92	91	94
Georgia	97 <sup>2</sup>	97 <sup>2</sup>	98 <sup>2</sup>
Hawaii	83 <sup>3</sup>	—	—
Idaho	97 <sup>3</sup>	91 <sup>3</sup>	96 <sup>3</sup>
Illinois	98 <sup>2</sup>	93 <sup>2</sup>	100 <sup>2</sup>
Indiana	96	94	97
Iowa	95	—	—
Kansas	93	90	94
Kentucky	97	96	99
Louisiana	91 <sup>2</sup>	87 <sup>2</sup>	93 <sup>2</sup>
Maine	90 <sup>2</sup>	91 <sup>2</sup>	91 <sup>2</sup>
Maryland	75	58	85
Massachusetts	93 <sup>3</sup>	88 <sup>3</sup>	94 <sup>3</sup>
Michigan	94 <sup>3</sup>	92 <sup>3</sup>	93 <sup>3</sup>
Minnesota	97	96	97
Mississippi	93	88	95
Missouri	96 <sup>2</sup>	90 <sup>2</sup>	97 <sup>2</sup>
Montana	99 <sup>3</sup>	99 <sup>3</sup>	99 <sup>3</sup>
Nebraska	91 <sup>2</sup>	90 <sup>2</sup>	92 <sup>2</sup>
Nevada	71	65	77
New Hampshire	73 <sup>2</sup>	69 <sup>2</sup>	73 <sup>2</sup>
New Jersey	94	86	96
New Mexico	78	75	88
New York	92 <sup>2</sup>	81 <sup>2</sup>	97 <sup>2</sup>
North Carolina	85 <sup>2</sup>	82 <sup>2</sup>	87 <sup>2</sup>
North Dakota	77	83	73
Ohio	93	85	96
Oklahoma	99	98	99
Oregon	90	89	—
Pennsylvania	97 <sup>2,3</sup>	92 <sup>2,3</sup>	99 <sup>2,3</sup>
Rhode Island	74 <sup>2</sup>	77 <sup>2</sup>	73 <sup>2</sup>
South Carolina	75 <sup>2</sup>	68 <sup>2</sup>	79 <sup>2</sup>
South Dakota	93	91	93
Tennessee	80	74	81
Texas	95	94	95
Utah	69 <sup>2</sup>	65 <sup>2</sup>	73 <sup>2</sup>
Vermont	82 <sup>2</sup>	82 <sup>2</sup>	83 <sup>2</sup>
Virginia	95 <sup>2</sup>	92 <sup>2</sup>	97 <sup>2</sup>
Washington	99 <sup>2</sup>	99 <sup>2</sup>	99 <sup>2</sup>
West Virginia	96 <sup>2</sup>	97 <sup>2</sup>	95 <sup>2</sup>
Wisconsin	98 <sup>2,3</sup>	96 <sup>2,3</sup>	99 <sup>2,3</sup>
Wyoming	49 <sup>4</sup>	49 <sup>4</sup>	49 <sup>4</sup>

Notes: A dash (—) indicates data were not available. Percentages have been rounded to the nearest whole number.

<sup>1</sup>A majority of states provided data for 2004-05, but a significant number of states were only able to provide information for the 2003-04 school year at the time of publication.

<sup>2</sup>Data represent the 2003-04 school year. Nebraska's data represent the 2002-03 school year.

<sup>3</sup>State does not collect "highly qualified" teacher data by class. Instead, data represent the percent of teachers or full-time equivalencies (FTEs) designated as "highly qualified" by the state.

<sup>4</sup>Wyoming is in the process of refining its "highly objective uniform state standard of evaluation" (HOUSSE) procedures for veteran teachers. Therefore, these figures reflect data as of October 2005 and are not final.

SOURCE: Editorial Projects in Education Research Center, 2005

CONTINUED FROM PAGE S6

Education Department's handling of the NCLB law's teacher-quality provisions, such as Ross Wiener, the policy director of the Education Trust, a Washington-based advocacy group for poor and minority students. He said that the threat of punishments for not meeting the deadline had helped create an "unfortunate dynamic" that made honest and accurate information a liability.

"You'd hope honest information would lead to talks with the legislature and higher education institutions, with state school officials saying here's what we need to bridge the gaps," Mr. Wiener said.

Instead, low numbers tended to get the wrong kind of attention from the news media and lawmakers, he said. And once a state had posted high numbers, a downward revision was hard, he added.

Mr. Rheault, the Nevada superintendent, said that legislators in his state had wanted explanations for the low teacher-quality numbers. Seventy-one percent of Nevada's classes overall, and just 65 percent of its classes in high-poverty schools, were taught by highly qualified teachers last year, according to the state's figures. Those proportions were up slightly over the previous two years.

With more than two-thirds of Nevada's new teachers coming from out of state, districts have problems making sure the recruits have passed the state's multisubject test required for highly qualified status, Mr. Rheault said.

In Delaware, education officials don't expect to be able to produce teacher-quality figures for classes until the end of this school year.

"We have tried to meet all the reporting requirements we can," said Robin Taylor, the associate superintendent for assessment and accountability. "But we have limited resources in terms of [money] and personnel and programming."

Initially, some Delaware districts told the state that 100 percent of their classes were taught by highly qualified teachers. "I knew that wasn't right," Ms. Taylor said. The state now has a process for verifying such data.

### Complicated Effects

Education Department officials cited four states' teacher-quality data systems as particularly sound: Colorado, Georgia, Kentucky, and Ohio.

In Ohio, state officials said they benefited from an extensive data system already in place. "We were able to build the high-quality-teacher data on top of that," said Marilyn B. Troyer, the associate superintendent for teacher quality. Ohio reported that 82 percent of classes overall and 78 percent of classes in high-poverty schools were taught by highly qualified teachers in the 2002-03 school year. Those proportions had risen to 93 percent and 85 percent, respectively, last year. "I still don't anticipate reaching 100 percent by the end of this school year," she said.

She said she thought that the state's numbers had been helped by schools' greater care in assigning teachers to classes and by new professional-development opportunities that had helped teachers win highly qualified status. A just-launched pilot program that offers bonuses to some teachers working in shortage fields in high-poverty schools might help shrink Ohio's teacher-quality gap, Ms. Troyer said.

Across the country this year, districts are grappling with the effects of the nearly 4-year-old federal law. They are also seeing its complications play out

in schools and classrooms.

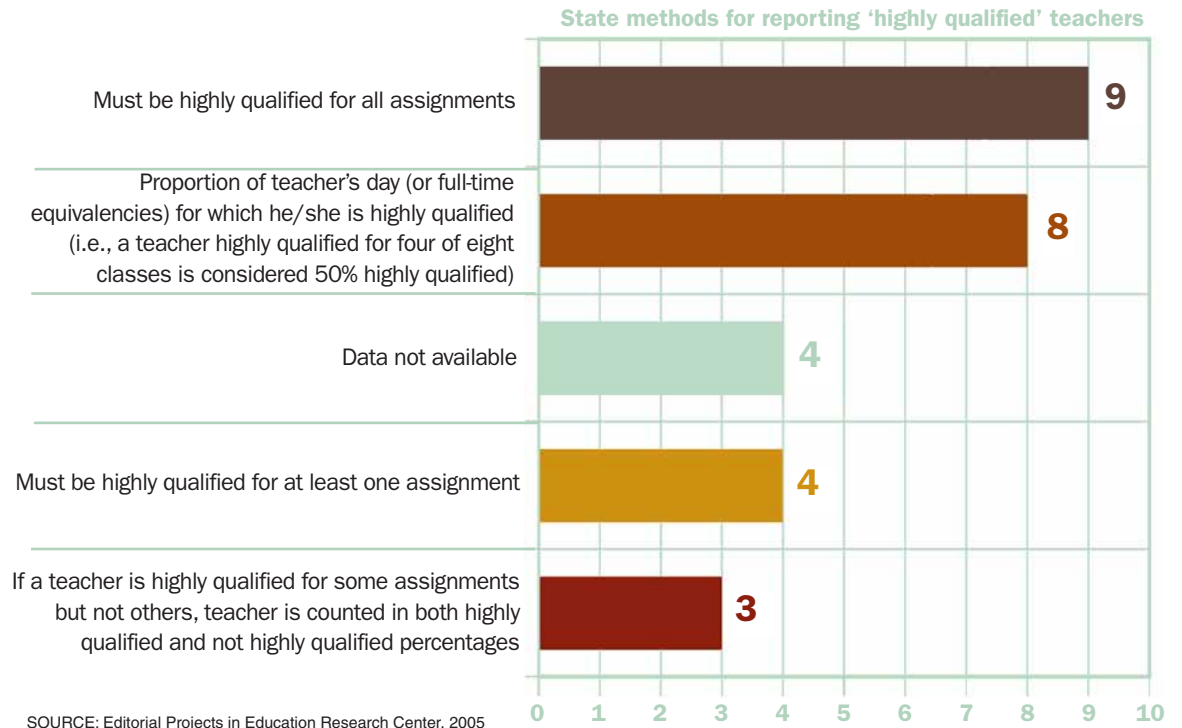
Stephen C. Lewis, the director of human resources for the 12,000-student Gresham-Barlow district near Portland, Ore., said that district administrators met this fall with all 19 principals, pinpointing for them where each of their faculty members stood in meeting the teacher requirements. Of the 660 teachers, fewer than 50 were still in the process of getting highly qualified status—most at the middle school level, according to Mr. Lewis. But a handful had yet to get in their initial paperwork so that their status could be officially determined, he said.

Because of the challenges the law poses for middle schools, a teaching “block” combining English and history might be done away with in the Gresham-Barlow system. The block arrangement requires teachers to be highly qualified in both subjects—a difficult task given that many hold elementary certification, which is not enough to show subject mastery in either field.

Mr. Lewis said he suspected that some teachers who were taking courses to gain highly qualified status knew more than the college instructors who taught them. But he cited the district’s small high school for students who weren’t fitting in elsewhere as a success story. “We had to work really hard to get highly qualified staff there,” he said, because typically, the teachers are responsible for more than one subject. As a result of that effort, he said, “I’m sure the faculty is better.” ■

### Who is ‘Highly Qualified’?

The No Child Left Behind Act requires states to report the percent of classes taught by “highly qualified” teachers, perhaps the best measure since teachers can satisfy that requirement for some subjects they teach but not others. Twenty-eight states, as shown below, however, report the percent of highly qualified teachers more generally. Nine of those 28 states report only the more general measure and not whether teachers are highly qualified for each of their assignments or classes.

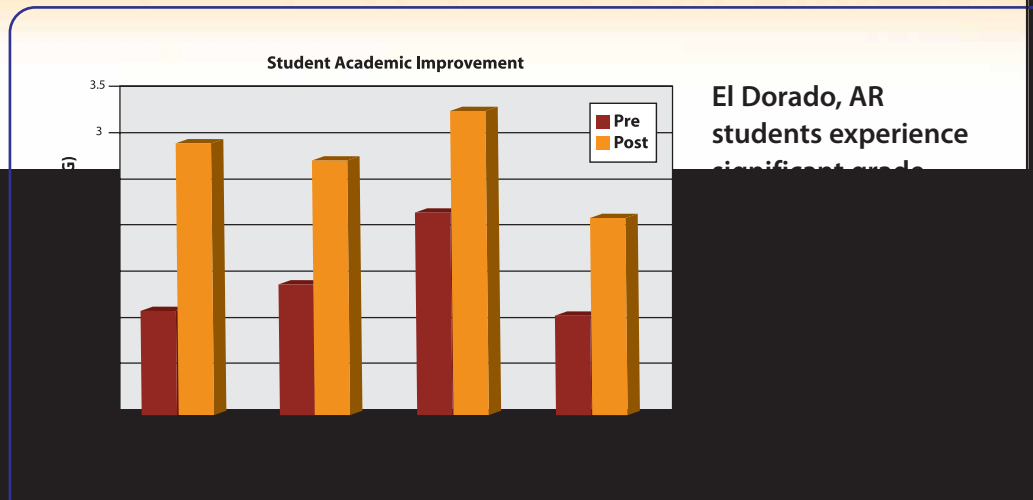


# Remove the Risk from “At Risk” Schools

## Prescriptive Instruction

Automated Differentiated Instruction

- Automated intensive intervention and remediation program, invaluable for At Risk students
- K-5 Reading and Math curriculum



Allow us to help you solve your educational challenges with a FREE phone or on-site consultation. Call 1-888-665-8307 or visit [prescriptiveinstruction.com](http://prescriptiveinstruction.com)



# States Confront Definition of 'Proficient' As They Set the Bar for Lots of New Tests

By Lynn Olson

As states add reading and math exams in previously untested grades to comply with the No Child Left Behind Act, they will have to determine the level of performance considered "proficient."

In particular, states must figure out how to make their achievement standards on the new tests mesh with those in the grades already being tested, so that the progression of growth expectations across grade levels is smooth. Otherwise, 4th graders who are rated proficient in mathematics one year may suddenly score below that level the next simply because the standard, or cut point, has shifted.

"I think it's causing some difficulties," Robert L. Linn, a professor of education emeritus at the University of Colorado at Boulder, said of state efforts to set performance standards.

A survey conducted this fall by the Editorial Projects in Education Research Center found that at least 11 states set new achievement levels in reading/language arts in the 2004-05 school year. About nine states did so in mathematics.

Those numbers are expected to grow substantially this school year, as nearly half the states administer reading and math tests in more grades.

The federal law requires states to give tests in reading and math in grades 3-8 and at least once

in high school, starting this school year.

The combination of new performance standards and tests will make it even harder to determine if schools are really improving, based on whether they have made adequate yearly progress under the federal law.

As states add performance standards or revise the scores students need to qualify as "proficient," it may be unclear if the bar has been raised, lowered, or kept largely the same.

## Comparisons Difficult

Figuring out the height of the bar is not easy, responses to the EPE Research Center survey show.

In Arizona, for example, officials held a series of meetings last May to set new achievement levels in reading, math, and writing for their state's tests in grades 3-8 and high school. State officials report that high school students now must answer a lower percentage of items correctly to meet the proficiency standard. But the tests also now contain more items, so students must show more knowledge to be rated proficient.

Arkansas also set new performance levels in 2005 for its reading and math exams. "On balance, the cut scores are generally comparable," a state education department official reported, "although at one particular grade or another, the cut score may be somewhat higher or lower."

"It is difficult to make an exact comparison," the Arkansas official continued, in response to the EPE Research Center survey, "since the content standards being measured have been revised and since the design of the literacy portion of the examination has changed."

What states want to avoid, said Scott Marion, a vice president of the Center for Assessment, a Dover, N.H.-based group that works with states to improve their testing-and-accountability systems, are erratic swings in performance from grade to grade because of where they've set the bar.

"If you have an assessment in grades 4, 8, and

11 and now you're going to fill in the rest of the grades, do you go back and completely revisit all your performance standards, which some folks are doing," Mr. Marion said, "or do you try and set new standards for the new tests and live with your old ones where they were?"

## Seeking Consistency

Some states, such as Arizona, have developed a single "vertical" scale that summarizes student achievement across grade levels, at least in grades 3-8.

Such scales, according to Robert W. Lissitz, a professor of education at the University of Maryland College Park, assume that tests at different grade levels focus on similar math or reading concepts even though they measure different content. Students are expected to improve on the scale each year as their math or reading skills increase.

But Mr. Lissitz and other assessment experts say that vertical scales are hard to construct and are based on questionable assumptions about how common the content really is across grades.

He and others advocate what they call "vertically articulated" or "vertically moderated" standards. Such methods rely on a combination of human judgment and statistical analyses. They consider both the content standards and test difficulty in each grade, along with data on how students actually perform, to set cutoff scores.

The assumption, said Mr. Marion of the Center for Assessment, is that if 50 percent of a state's 3rd graders are proficient in mathematics, "and you don't think 4th grade math is all that different, your best guess is 50 percent of the kids should be proficient in grade 4, too."

"That's not deterministic," he said. "It allows you to set a starting point."

## 'A Purposeful Act'

In Michigan, for example, curriculum standards were revised in 2004, based on the NCLB testing requirements, so that grade-level expectations are now more rigorous and specific. The state not only added new tests in grades 3, 5, 6, 7, and 8, but also shifted from a spring to a fall testing date.

The state plans to set performance standards and cutoff scores for the new tests in late December and early January. As one step in that process, said Ed Roeber, the state's testing director, committees will review books in which test items are arranged in order of difficulty and determine where to set the proficiency bar.

Those books also will show where that bar would be placed to maintain a level of proficiency consistent with that in adjacent grades. What the committees decide from there is really unconstrained, Mr. Roeber said.

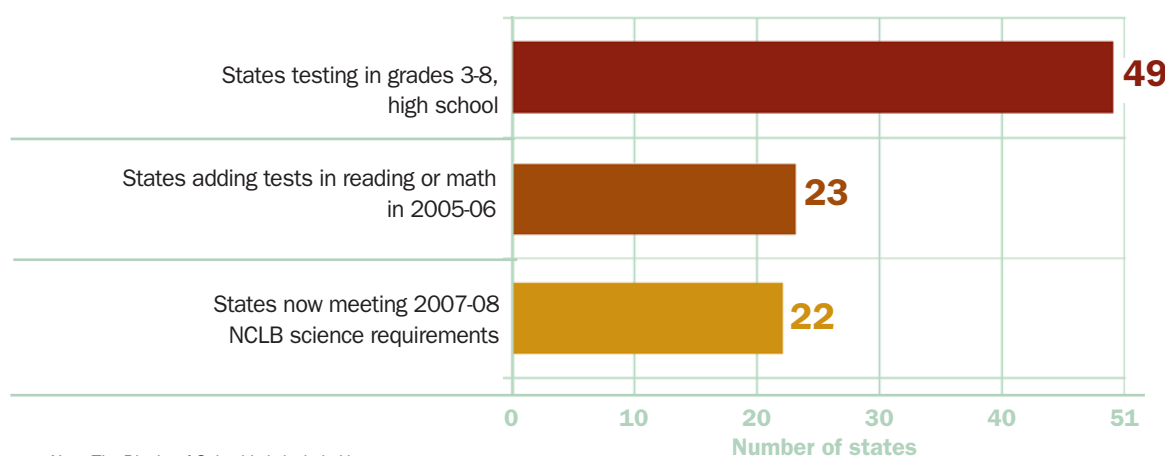
"Even at the grades where we've had tests, the committees could set standards higher or lower," he said. "We don't want them to do it by accident; we want it to be a purposeful act."

For now, said Mr. Lissitz, "nobody has a real solid answer" on the best method for making such judgments.

"It's a hard thing, because the models that we have are being developed as we speak," he said. "Right now, we have answers, but they're not as satisfactory as they will be in a couple of years." ■

## Meeting Requirements

States are moving to meet the testing requirements of the No Child Left Behind Act. Almost all states will test students in reading and math in grades 3-8 and once in grades 10-12 in 2005-06. Slightly fewer than half of all states currently have standards-based science tests in each of three grade spans: 3-5, 6-9, and 10-12. The nearly 4-year-old federal law requires states to have such science tests in place by 2007-08.



Note: The District of Columbia is included in this analysis. Total state count = 51.  
SOURCE: Editorial Projects in Education Research Center, 2005

## Special Report: Testing Takes Off

## State Test Programs Mushroom as NCLB Mandate Kicks In

Nearly half of states are expanding tests into more grades in 2005-06 school year

BY LYNN OLSON

You could describe it as the year of the tests. Twenty-three states are expanding their testing programs to additional grades this school year to comply with the federal No Child Left Behind Act.

"Every group I've been talking to, I've just said, 'Be patient with us this year,'" said Alexa E. Posny, the deputy commissioner of education in Kansas, which is adding reading and math tests in four grades in each subject.

"First, it's the sheer volume," she said. "In the past, we would develop 4,000 test items; we're developing 18,000 items. Second is the number of teachers who have never participated in state assessments, so it's a whole new ballgame for them. And then there's the overwhelming amount of data that will be available because there are so many more grades."

Forty-eight states and the District of Columbia will give standards-based tests in reading and mathematics in grades 3-8 and at least once in high school this school year, as required by the nearly 4-year-old federal law, according to a survey by the Editorial Projects in Education Research Center.

The holdouts are Iowa and Nebraska. Districts in Iowa give the Iowa Tests of Basic Skills, a national test not designed to measure state or local content standards, while districts in Nebraska craft their own tests, except for a state writing exam.

In devising the new tests, most states have defied predictions and chosen to go beyond multiple-choice items, by including questions that ask students to construct their own responses.

But many state officials are

worried that, over the long run, such decisions could push their assessment programs into the red. In addition, despite efforts aimed at getting test results back faster, they fear that the sheer volume of assessments could generate delays and errors in releasing scores.

But many state officials are

worried that, over the long run, such decisions could push their assessment programs into the red. In addition, despite efforts aimed at getting test results back faster, they fear that the sheer volume of assessments could generate delays and errors in releasing scores.



Empty testing envelopes sit at a contractor's warehouse in Dover, N.H.

States have generally filled in the gaps in their testing programs with assessments that mirror those in other grades. Many analysts had predicted that, because of costs, states would rely solely on multiple-choice tests for grades and subjects in which they had not tested previously.

worried that, over the long run, such decisions could push their assessment programs into the red. In addition, despite efforts aimed at getting test results back faster, they fear that the sheer volume of assessments could generate delays and errors in releasing scores.

In general, states have filled in

jects in which they had not tested previously.

A 2003 study by the investigative arm of Congress estimated that it would cost states \$1.9 billion to meet the testing requirements over the six years of the federal law's authorization if they relied solely on multiple-choice questions that could be machine-

scored. But it would cost states \$3.9 billion if they used a mix of multiple-choice and open-ended items, and up to \$5.3 billion if the tests required hand-scored, written responses, according to the agency, now called the Government Accountability Office.

"We're developing tests that have the same format, blueprint, rigor, as the tests that we already have in place," said Jeffrey Nellhaus, the deputy commissioner in the Massachusetts Department of Education. The reading and math tests require students to construct their own responses to some items, in addition to answering multiple-choice questions.

The Massachusetts program is expected to cost "somewhere in the order of \$10 million to \$12 million a year," Mr. Nellhaus estimated, compared with the just over \$7.6 million the state is receiving from the federal government this year to cover such costs.

### Beyond Multiple Choice

In Nevada, which added reading and math tests in grades 4, 6, and 7 this year, the money has so far been adequate to create tests with a mix of multiple-choice and constructed-response questions, said Paul M. La Marca, the state's assistant deputy superintendent.

"We know that tests are driving curriculum, to some extent, so we think it's important to have items that stretch the cognitive demand of the students," he said. "You can do that with multiple-choice items, but you have a better chance of doing it with other types of items."

Still, Mr. La Marca said, the state doesn't have as many constructed-response items as it might want. "The balance of the tests is skewed toward multiple-

choice," he said. "It's more than just a cost issue. In our state, we have significant pressure for quick turnaround time, so that almost hamstring us a little bit."

New Jersey officials announced Nov. 16 that, in addition to the tests the state already gives in grades 3, 4, 8, and high school, it would add a commercial test aligned to its reading and math standards in grades 5, 6, and 7 for the 2005-06 school year as an interim measure, while it works to revise its entire testing program to provide better diagnostic information for educators.

"Ultimately, we want to build a more robust, more rigorous state system, hopefully to incorporate a performance assessment which we have been piloting for the last three years," said Acting Commissioner of Education Lucille E. Davy. She said the state worked closely with the U.S. Department of Education to ensure the plan could meet federal requirements, although the system still has to go through a peer-review process before it can be approved, as is true for all states.

In Mississippi, meanwhile, officials decided to drop all short-answer questions in the 2005-06 school year to facilitate speedier scoring of test results. It now has only multiple-choice items, except for a state writing test.

Kansas has suspended the use of all items that require an extended response from students and can't be machine-scored.

"It doesn't mean in the future we may not add those items," Ms. Posny said. "We want to figure out how we can do that and still ... score [the test] online."

Fifteen states in total will rely solely on multiple-choice items to measure student knowledge, with the exception of their writing

## Federal Review Puts State Tests Under Scrutiny

BY LYNN OLSON

By the end of this month, 26 states will have undergone a "peer review" to determine whether their standards and tests meet the requirements of the federal No Child Left Behind Act.

The reviews, conducted by a team of at least three experts in the fields of standards and assessment, are required under the law. The reviewers do not look at the standards and tests themselves, but at documents showing that the assessments meet the law's requirements.

As of Nov. 21, the U.S. Department of Education had posted letters to six states—Maryland,

North Carolina, South Carolina, South Dakota, Texas, and West Virginia—on its Web site granting them "deferred" or "final review pending" status under the law.

To receive deferred approval, a state must be able to fully implement its standards and tests this school year; "final review pending" indicates the state still has not met the preponderance of NCLB testing requirements and must submit more evidence. Such documentation can range from technical reports or test manuals, to state statutes and regulations, to memos summarizing the testing program.

One of the issues giving states trouble is a requirement to provide "performance descriptors" that explain the competencies a student must master in mathematics or reading to reach a particular performance level, such as "proficient."

Those descriptions must pertain to specific academic content,

said Sue Rigney, an education specialist at the Education Department. "What's not acceptable is to see these very generic descriptors that are the same across grade levels and content areas," she said.

States also are struggling to prove the quality of their alternate assessments for students with disabilities who cannot take the regular exams, and those tests' link to state standards.

States were required to have alternate assessments in place by 2001 under the prior reauthorization of the Elementary and Secondary Education Act. But it wasn't until last summer that the department provided guidance about the criteria for such tests if they are pegged to other than a grade-level standard.

As a result, said Rachel Quenemoen, a senior research fellow at the National Center on Education Outcomes, based at the University of Minnesota-Twin Cities,

states have been designing alternate assessments during a period of constantly changing policy and emerging research.

"The states have come a long, long way," she said, "but the depth of the research and the attention that it's gotten is very, very slim. The conditions under which states were working with alternate assessments have changed dramatically."

### 'Really Encouraged'

As they add tests this school year to comply with the nearly 4-year-old NCLB law, states must design performance standards for those new tests that mesh with those already set for other grades.

In many cases, that will require states to revisit their existing cut-off scores, so that students who perform well in one grade can reasonably be expected to perform well in the next.



Boxes of tests arrive in the scanning room

Much of that standards-setting will occur over the summer of 2006, after the new tests are given for the first time this coming spring. That will require the federal Education Department to gather additional evidence from most states before it can give full approval to their systems.

Even so, said Kerri L. Briggs, a senior policy adviser in the office of the deputy U.S. secretary of education, "I think we're really encouraged at this point about where states are."

Fewer than a dozen states, she

## Special Report: Testing Takes Off

### States Adding Tests in 2005-06

Grades:	READING								MATH									
	3	4	5	6	7	8	3	4	5	6	7	8	3	4	5	6	7	8
Connecticut	•	•	•				•	•	•									
Illinois		•	•	•					•	•	•				•	•	•	
Kansas	•	•	•	•			•	•	•	•	•							
Kentucky	•	•	•	•			•	•	•	•	•							
Maine	•	•	•	•			•	•	•	•	•							
Massachusetts		•	•	•			•	•	•									
Michigan	•	•	•	•			•	•	•	•	•							
Minnesota		•	•	•					•	•	•							
Missouri	•	•	•	•			•	•	•	•	•							
Montana	•	•	•	•			•	•	•	•	•							
Nevada		•	•	•					•	•	•							
New Hampshire	•	•	•	•	•		•	•	•	•	•							

SOURCE: Editorial Projects in Education Research Center

FOOTNOTES:

- 1 The District of Columbia previously used the SAT-9 in grades 1-11. In 2005-06, it will use a standards-based exam in grades 3-8 and 10.
- 2 Louisiana previously administered the Iowa Test of Basic Skills. In 2005-06, it will introduce an assessment program that combines the ITBS and criterion-referenced items, in grades 3, 5, 6, 7, and 9.
- 3 New Hampshire, Rhode Island, and Vermont suspended testing for 2004-05. These states have implemented jointly developed assessments in grades 3-8 in 2005-06.

### Testing Changes At a Glance

	NUMBER OF STATES
States testing in grades 3-8, high school	48*
States adding tests in reading or math	23
States with all multiple-choice except for writing	15

\*Does not include the District of Columbia. SOURCE: EPE Research Center

tests, according to the EPE Research Center survey: Arizona, California, Georgia, Idaho, Iowa, Kansas, Mississippi, North Carolina, Oklahoma, Oregon, South Dakota, Tennessee, Texas, Virginia, and Utah.

The District of Columbia this year switched from a multiple-choice test that was not aligned with its academic-content standards to a new set of standards and tests based on those given in Massachusetts.

#### Costs a Concern

Although many states said the federal government has provided enough money to cover the development of new tests, they worry about the costs in future years.

Washington state's tests include a 50-50 mix of multiple-choice and open-ended questions. While federal funding has so far been sufficient to expand the tests to grades 3, 5, 6, and 8, said Greg B. Hall, the state's assistant superintendent for assessment and research, starting next fiscal year the state will run a deficit in its testing budget that is expected to increase over time.

CONTINUED ON PAGE 12

### Reading Programs That Change Lives

Our classroom-proven CARS and STARS programs make a difference every day, and this fall they are better than ever with exciting new content—all for the same low price! Also new this fall is FOCUS, the perfect program to dovetail with our CARS and STARS series. For students needing more targeted practice, the FOCUS series offers strategy-specific practice books written at levels 1-8.

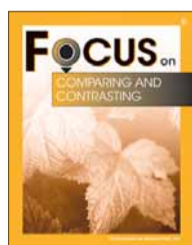
The programs:

- Meet state standards
- Are research-based—download the research papers for FREE from our Web site
- Easily differentiate reading instruction
- Are guaranteed to raise student performance—and test scores

For more information and to order your FREE samples visit [www.CurriculumAssociates.com/Reading800](http://www.CurriculumAssociates.com/Reading800) or call 1-800-225-0248



▲ NEW Copyrights:  
SAME LOW COST PER STUDENT!



▲ NEW 48-Book Series!

FROM THE PUBLISHER OF THE:

CARS® and STARS™ • TEST READY® • BRIGANCE® • WRITE! • QUICK-WORD® Programs

As states undergo federal "peer review" of their standards and testing systems under the No Child Left Behind Act, the U.S. Department of Education says it will insist that all states comply with the law.

at Dover, N.H.-based Measured Progress.

noted, had received final approval of their testing systems under the previous reauthorization of the federal education law, in 1994. This time around, the department has made it clear that waivers of the law's testing requirements will not be acceptable.

"From the get-go, we've been really serious about this provision, in particular," said Ms. Briggs, "and we have every intention of implementing it."

## Special Report: Testing Takes Off

# As States Add Tests, Concerns Rise Over Costs and Logistical Problems

CONTINUED FROM PAGE 11

The reason, he said, is the cost involved in scoring so many additional open-ended items, "which we haven't had to do yet."

Rhode Island, New Hampshire, and Vermont have jointly developed grade 3-8 reading and math tests to meet the federal law's requirements. While federal aid has covered those costs so far, "we're all worried about what will happen when this money goes away," said Mary Ann Snider, the director of assessment and accountability for the Rhode Island Department of Education.

In August, Connecticut became the first and, so far, only state to sue the federal government over the No Child Left Behind law, charging that federal funding falls short of what is needed to meet the law's requirements.

Connecticut officials have sought unsuccessfully to get out of expanding their testing in core subjects beyond grades 4, 6, and 8. An estimate by the state education department pegs the cost of providing tests in the additional grades required under the NCLB law at \$41.6 million by

2008, compared with \$33.6 million that the state is slated to receive from the federal government by then for testing. (See *Education Week*, Aug. 31, 2005.)

The annual testing requirement is a linchpin of the federal law. Schools and districts are required to meet annual performance targets for their student populations as a whole and for certain subgroups of students. Those that receive federal Title I money and that fail to meet their targets for two or more years face penalties.

## Earlier Testing Dates

At least some states are shifting their testing dates to try to get results back sooner. New Hampshire, Rhode Island, and Vermont started giving their new jointly developed tests in the fall, rather than in the spring.

That timing was intended, in part, to ensure that schools would know by the winter whether they have met their performance targets under the federal law and would be subject to any of the law's sanctions, Ms. Snider said.

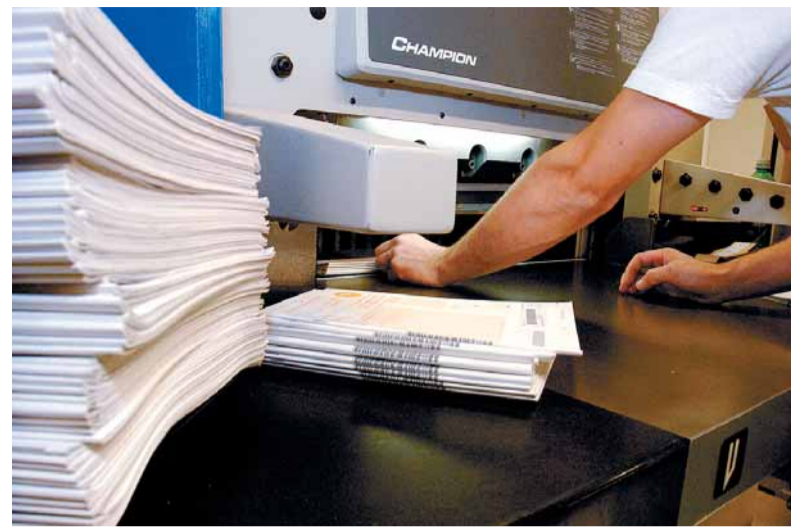
Michigan shifted from January to October testing, but not with-

out some mishaps. "The idea is to give data back to teachers while they still have the students, and they still have the energy to do something about the results," said Edward D. Roeber, who directs the state testing program.

Still, the change "cut out four months in the test-preparation schedule, which made the summer very challenging," he said. And when the state's test contractor, Pearson Education Inc., failed to get enough tests delivered to school districts on time, the state was forced to extend the testing window by two weeks.

Other states have made more minor adjustments. Kansas, for instance, moved its testing dates up by one week, to March 1. It also has put its writing, science, history, and government tests on hold for a year, while it gets the new reading and math exams in place.

Massachusetts consolidated all of its reading tests in April, so that they would not compete with tests in other subjects during May. New York state adopted a flexible schedule for the statewide administration of its grades 3-8 exams for this school year, to permit schools to give one session per day



A worker inserts bundles of tests into a binding-removal machine.

State officials anticipate that it will take longer to report results this school year, as they try to craft new performance standards in additional grades and cope with the extensive amounts of data. Another concern is that the sheer volume of tests will increase the likelihood of errors in everything from the production of test booklets to the processing and reporting of results.

at each grade level. Maine has opted to replace an existing state test with the SAT college-admissions exam at the high school level, starting next spring.

Even so, state officials generally anticipate that it will take longer to report results this school year, as they try to craft new performance standards in the additional grades and cope with the extensive amounts of data.

"It will be delayed, there is no doubt in my mind," said Ms. Posny of Kansas.

## Potential for Error

One concern is that the sheer volume of tests will, inevitably, create logistical problems for schools and an increase in administration and scoring errors.

"The more you add to the test contractors' plate and the quicker they have to report the results back, the probability goes up that errors are going to be made," said George Madaus, a professor emeritus of education at Boston College who co-wrote a May 2003 report on the widespread errors in standardized tests. "You're stretching the capacity of a limited number of companies that do this work."

Among the states that have suffered from scoring glitches is Nevada. In 2002, the state board of education required the San Antonio-based Harcourt Educational Measurement to pay penalties totaling \$425,000 because of a mistake that threw off the scores of nearly 31,000 students who had taken the state's high school exit exam in math.

"We're very worried about accuracy from the vendor because we have, unfortunately, been snake-bitten," Mr. La Marca said.

In Michigan, which is giving 216 different test forms in grades

3-8 this year, state officials are doing "a lot of extra checking," said Mr. Roeber.

"There are just more chances to screw up," he said. "My staff and I have been working 18 hours a day and some weekends."

And state officials aren't alone. "School districts are kind of reeling," Mr. Roeber said.

One issue for schools is simply having enough staff members to administer the tests and to provide accommodations, such as more time, for all the students who need them because of disabilities or limited English skills.

"As with most states, we have an extensive list of accommodations, which is great for students but puts the school staff in a difficult position, because many of the accommodations require alternate settings and additional staff," said Tim Kurtz, the director of assessments for the New Hampshire Department of Education. "Every single adult human being is involved in testing somehow."

At least one thing is clear: With more tests, in more grades, soon many more teachers will be focused on test results.

"The existing evidence suggests that when a grade level is tested, teachers pay a lot more attention to what's on the test," said Joan L. Herman, the co-director of the Center for Research on Evaluation, Standards, and Student Testing, or CRESST, at the University of California, Los Angeles.

"So as more grade levels are being tested," she said, "you can now expect that every teacher will be paying attention to what's on the test and, in the best case, aligning their instruction with standards and, in the worst case, engaging kids in a curriculum of test preparation."

Shorten the distance between dreaming of an advanced degree and earning it.

## Fischler School of Education and Human Services offers

- Student-focused national faculty
- More than 60 specializations available
- Powerful networking opportunities
- Field-based and online delivery systems

## Degree programs on-site, online, worldwide

- Associate of Arts in Early Childhood Education (A.A.)
- Bachelor of Science in Education (B.S.)
- Master of Science in Education (M.S.)
- Educational Specialist (Ed.S.)
- Doctor of Education in Child, Youth and Human Services (Ed.D.)
- Doctor of Education in Education Leadership (Ed.D.)
- Doctor of Education in Instructional Technology and Distance Education (Ed.D.)
- Doctor of Education in Organizational Leadership (Ed.D.)

Contact us to learn more about your opportunities with FSEHS  
800-986-3223  
[www.SchoolofEd.nova.edu](http://www.SchoolofEd.nova.edu)



## Special Report: Testing Takes Off

# Benchmark Assessments Offer Regular Checkups On Student Achievement

BY LYNN OLSON

School districts worried about how students will perform on end-of-the-year state tests are increasingly administering “benchmark assessments” throughout the year to measure students’ progress and provide teachers with data about how to adjust instruction.

Nearly seven in 10 superintendents surveyed for *Education Week* this past summer said they periodically give districtwide tests, and another 10 percent said they planned to do so this school year. Such tests typically are aligned to state or district standards for academic content and given three to five times during the year. Some are given as often as monthly.

Most benchmark assessments take one hour each for reading and mathematics, but may include other subjects. Extensive reporting systems break down test results by the same student categories required under the federal No Child Left Behind Act, such as by race, income, disability, and English proficiency, in addition to providing individual progress reports at the district, school, classroom, and student levels.

“I do believe that three years from now, certainly five years from now, no one will remember a time when there weren’t benchmarks,” said Robert E. Slavin, the director of the Center for Data-Driven Reform in Education, at Johns Hopkins University.

## Burgeoning Market

That’s certainly what test vendors hope. Last year, Eduventures Inc., a market-research firm based in Boston, identified benchmark assessments as one of two high-growth areas in the assessment industry, alongside state exams, with a compound annual growth rate of greater than 15 percent. The company predicted that by 2006, what it called “the formative-assessment market”—using a term sometimes treated as a synonym for benchmark assessment—would generate \$323 million in annual revenues for vendors.

But while many assessment experts agree that the idea of frequent testing of students to monitor their learning and adjust instruction is sound, some also warn that districts should take a close look at what they’re getting for their money and how they are using such exams.

“You might say that the message here is, ‘Get a second opinion,’” said Grant Wiggins, the president of Authentic Education, a Hopewell, N.J.-based consulting service that works with districts.

It’s no secret why districts are turning to benchmark tests. The

No Child Left Behind Act, signed into law by President Bush in January 2002, and states’ own accountability systems have created a high-stakes environment in which both districts and schools can face penalties for failing to meet performance targets.

In this standards-based environment, the feeling is that the sooner and more often schools have information about how they’re doing against the standards, the better.

“The reason that there is a boom in benchmark assessments is that most states and school systems are providing nothing more than autopsy reports right now,” said Douglas B. Reeves, the founder of the Center for Performance Assessment, a private consulting organization based in Denver that works with districts to design fair and rigorous assessments and classroom activities. “They tell you why the patient died at the end of the year, and then marveled that the patient didn’t get any better.”

Studies by the Washington-based Council of the Great City Schools, the Austin, Texas-based National Center for Educational Accountability, and others have found that one feature of high-achieving districts is their use of periodic, benchmark assessments to track student achievement and make adjustments.

“Good formative assessments, good benchmark assessments,” Mr. Reeves said, “provide feedback throughout the year, and that is far more fair to principals and teachers, provided they are used wisely.”

## Vendors Vary

In the past few years, according to Eduventures’ 2004 report, “Testing in Flux,” new competitors have flooded the formative-assessment market, including:

- Major test publishers, such as the New York City-based CTE/McGraw-Hill and the San Antonio-based Harcourt Assessment;
- Test-preparation companies, including the New York City-based Princeton Review;
- For-profit providers that specialize in linking assessment results with prescribed remediation plans and curricula, such as the San Diego-based Compass Learning and the New York City-based Kaplan K-12 Learning Services;
- Nonprofit organizations, such as the Portland, Ore.-based Northwest Evaluation Association; and
- Suppliers of “whole-school-reform models,” such as the New York City-based Edison Schools Inc. and Mr. Slavin’s Baltimore-based Success for All Foundation,

CONTINUED ON PAGE 14



Silas Bender, a 3rd grader at London Towne Elementary School in Centreville, Va., takes a benchmark test.

Christopher Powers/Education Week

## Not All Teachers Keen on Periodic Tests

BY LYNN OLSON

John W. Hutcheson now teaches in a private Montessori school in Sammamish, Wash., after spending 25 years teaching in the Dallas school district. Looking back, he says the Texas district’s thrice-yearly benchmark assessments helped drive him out.

“The benchmarks themselves are a reflection of the standardized exams,” Mr. Hutcheson said, “which are only a small piece of learning. You progressively keep narrowing the curriculum down, so we end up preparing students for a world that doesn’t exist.”

Across the country, school districts are adopting benchmark assessments to help teachers modify instruction over the course of a school year. Yet many teachers remain wary. Like Mr. Hutcheson, they say their experience with such tests has been anything but positive.

In Philadelphia, a social studies teacher who asked not to be named said he found the use of benchmark assessments there “incredibly restricting and unrealistic.”

As part of a core high school curriculum, the 214,000-student school system uses a program involving multiple-choice tests given every six weeks, with immediate feedback to teachers and schools via a Web-based system of data analysis and reporting. The district describes the new standardized, college-preparatory curriculum and the related system of assessments as a critical element of its plan to improve secondary education. (See *Education Week*, Feb. 9, 2005.)

“Students found them totally meaningless and very intrusive, because it was another interruption, in addition to all

the other testing,” he said.

Mr. Hutcheson also complained of the time and stress associated with the tests used in Dallas. “We would spend entire afternoons analyzing benchmark results,” he said. “The district, every time the kids took the test, would print up a thorough record of how many answers they missed, the answers they put down, a list of subskills to be worked on, and a complete analysis of each test.”

Dallas school officials were unable to comment by press time.

Some districts have reported

Some classroom teachers see benchmark assessments as time-consuming intrusions that are exacerbating pressure to narrow the school curriculum to focus solely on subjects found on states’ standardized tests.

impressive results using similar methods.

When the Norfolk, Va., school district walked away with the \$500,000 Broad Prize in Urban Education this year, it was largely on the strength of its gains in reading and math scores and its progress in closing racial and ethnic achievement gaps. Officials there pointed to the strong focus on data-driven instruction as one key to the district’s success.

The 36,700-student district requires quarterly benchmark assessments in all grades. Ninety percent of Norfolk’s schools also have developed common assessments that teachers give monthly. And

teachers regularly meet in “data teams” to review the data, draw up common plans, and adjust instruction.

Over the past several years, the 12,000-student Santa Monica, Calif., school district has used a mix of teacher-designed tests and assessments linked to its adopted textbooks at the elementary school level. This year, secondary school teachers are meeting in departmental teams across sites to develop what the district is calling formative assessments in English, mathematics, science, and social studies that they’ll agree to give in common about three times a year.

“These are for teachers to really help guide their instruction,” said Maureen L. Bradford, the district’s director of educational services. “We feel like there probably isn’t something off the shelf that’s going to work for us; that teachers really need to come to one mind about what’s important to teach, and when to teach it and how to assess it appropriately. It’s a tremendous amount of work.”

Carol Jago, who chairs the English department at Santa Monica High School, praised the approach the school system is taking to developing the tests. “I hope we’re going to end up with essays or something that’s really authentic,” she said.

Still, Ms. Jago is worried.

“Inevitably, any time you try to institutionalize it, it becomes one more summative assessment that just happens before the state assessment,” she said, referring to a test given after teaching in the subject is completed. “So it’s right-headed, but I don’t think it’s something you can actually do properly because of the nature of the beast.”



## Special Report: Testing Takes Off

## Growing Demand for Benchmark Tests Fuels Policy Debate

CONTINUED FROM PAGE 13

which designed the 4Sight assessment series.

The products of such suppliers range from formatted tests linked to the standards in individual states, to item banks that districts and schools can use to develop their own assessments, to online testing, scoring, and reporting systems.

## Skimming the Surface?

Lorrie A. Shepard, the dean of the school of education at the University of Colorado at Boulder, voices caution about the trend.

While “not all formal benchmarking systems are bad,” she said, she worries about the effects of using 15- or 20-item multiple-choice tests that mirror the format of state exams to drive classroom instruction.

Previous research by Ms. Shepard and others has found that students who do well on one set of standardized tests do not perform as well on other measures of the same content, suggesting that they have not acquired a deep understanding.

“The data-driven-instruction fad means earlier and earlier versions of external tests being administered at quarterly or monthly intervals,” Ms. Shepard said. “The result is a long list of discrete skill deficiencies requiring inexperienced teachers to give 1,000 mini-lessons.”

Good benchmark assessments, she suggested, should include rich representations of the content students are expected to master, be connected to specific teaching units, provide clear and specific feedback to teachers so that they know how to help students improve, and discourage narrow test-preparation strategies.

Rather than trying to assess everything, added Mr. Reeves, the best benchmark tests focus on the most important state or district content standards. And they provide results almost immediately, in simple, easy-to-use formats, he said.

The National Center for Educational Accountability stresses that good benchmark assessments measure performance “on the entire curriculum at a deep level of understanding.” They also begin before grade 3 in both reading and math and provide a process to ensure that data on student performance are reviewed and acted upon by both districts and schools, the center says. In addition to such tests, it adds, districts may provide unit or weekly assessments that principals and teachers can use to monitor student progress.

## Approaches Differ

But in talking about benchmark assessments, not everyone means the same thing.

According to Mr. Slavin, some benchmark tests, like 4Sight, are

## What Are Benchmark Assessments?

While not everyone means the same thing by the term, benchmark assessments typically:

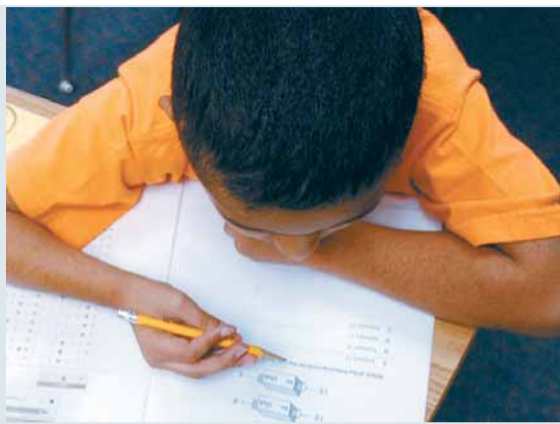
- Are given periodically, from three times a year to as often as once a month;

- Focus on reading and mathematics skills, taking about an hour per subject;

- Reflect state or district academic-content standards; and

- Measure students' progress through the curriculum and/or on material in state exams.

SOURCE: Education Week



Victoria Todd, a 3rd grader at London Towne Elementary, finishes one of 38 problems on a Benchmark Assessment Resource Tool test.

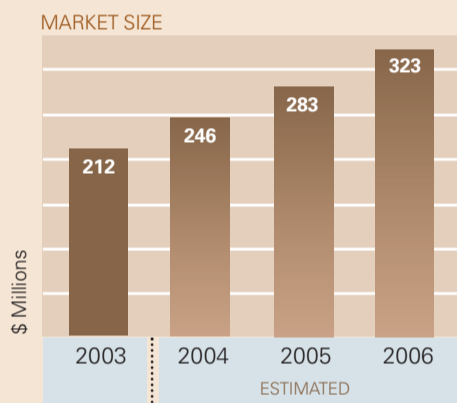
Christopher Powers/Education Week

Analysts see benchmark assessments as one of two high-growth areas in the testing industry, alongside state exams. Experts say good benchmark tests provide valuable feedback on students' mastery of key knowledge and skills. But some worry that as vendors have rushed in, quality has not kept pace.

## Benchmark-Test Market Foresees Growth

A 2004 report predicted that the market for benchmark or formative assessments would expand by a compound annual growth rate of more than 15 percent from 2003 to 2006.

SOURCE: Eduventures Inc.



## TEST MARKET

New competitors have emerged in recent years to supply school districts with benchmark assessments. They include:

**MAJOR TEST PUBLISHERS**, such as CTB/McGraw-Hill, based in New York City, and the San Antonio-based Harcourt Assessment;

**TEST-PREPARATION COMPANIES**, including the Princeton Review, based in New York City;

**SUPPLIERS** of whole-school-reform models, such as Edison Schools Inc., of New York, and the Success for All Foundation, of Baltimore.

**FOR-PROFIT PROVIDERS** that specialize in linking assessment results with prescribed remediation plans and curricula, such as the San Diego-based Compass Learning and the New York City-based Kaplan K-12 Learning Services;

**NONPROFIT ORGANIZATIONS**, such as the Northwest Evaluation Association, in Portland, Ore.

SOURCE: Eduventures Inc., Education Week

use them appropriately.

“Now we’re putting individual items in the hands of teachers,” he said, “saying, ‘You construct the test; make it as long or as short as you want.’ Do we think they have the understanding to know how much stock they can put in the generalizations they make from such exams?”

Some also worry that as vendors have rushed in, quality has not kept pace. The Eduventures report noted that many vendors have marketed formative assessments “on the basis of the quantity of exam items, as opposed to those items’ quality.” For example, companies may tout having tens of thousands of exam items, it said, although many of the items have not been extensively

field-tested or undergone a rigorous psychometric review.

“I think vendors in our space have found it challenging,” said Marissa A. Larsen, the senior product manager for assessment at the Bloomington, Minn.-based Plato Learning Inc., whose eduTest online assessment system is now used in more than 3,000 schools.

While districts sometimes apply the same psychometric standards to benchmark tests that are applied to high-stakes state exams, she said, “in many cases, that’s not what vendors in this space are trying to do. If we did that, it would be well beyond what districts could afford to buy for formative systems.”

Critics also say that even the

best benchmark assessments are more accurately described as “early warning” or “mini-summative” tests, rather than as true “formative” assessments, which are meant to help adjust teaching and learning as it’s occurring. In contrast, summative tests are designed to measure what students have learned after instruction on a subject is completed.

“Formative assessments are while you’re still teaching the topic, providing on-the-spot corrections,” said Mr. Kahl. “With benchmark assessments, you’re finished. You’ve moved on. Not that you don’t get individual student information, but at that stage, it’s remediation.”

## What Is ‘Formative’?

Yet Eric Bassett, the research director for Eduventures, said the terms formative and benchmark assessments are often used interchangeably in the commercial education market.

And that, some critics say, is precisely the problem.

“I recognize that I’ve lost the battle over the meaning of the term ‘formative assessment,’” said Dylan William, a senior researcher at the Educational Testing Service, based in Princeton, N.J.

In the 1990s, he wrote an influential review that found that improving the formative assessments teachers used dramatically boosted student achievement and motivation. Now that same evidence, he fears, is being used to support claims about the long-term benefits of benchmark assessments that have yet to be proven. “There’s a lack of intellectual honesty there,” Mr. William said. “We just don’t know if this stuff works.”

He and others say the money, time, and energy invested in benchmark assessments could divert attention from the more potent lever of changing what teachers do in classrooms each day, such as the types of questions they ask students and how they comment on students’ papers.

“If you’re looking, as you should be, at the full range of development that you want kids to engage in, you’re going to have to look at their work products, their compositions, their math problem-solving, their science and social-studies performance,” said Mr. Slavin of Johns Hopkins.

Mr. Wiggins of Authentic Education said that while some commercially produced benchmark assessments are far from ideal, they’re better than nothing. “I would rather see a district mobilizing people to analyze results more frequently,” he said. “That’s all to the good.”

The key point, he and others stress, is what use is made of the data.

“It’s only a diagnosis,” Mr. Slavin said. “If you don’t do anything about it, it’s like going to the doctor and getting all the lab tests, and not taking the drug.”