ESEA REAUTHORIZATION: OPTIONS FOR IMPROVING NCLB'S MEASURES OF PROGRESS

HEARING

BEFORE THE

COMMITTEE ON EDUCATION AND LABOR

U.S. HOUSE OF REPRESENTATIVES

ONE HUNDRED TENTH CONGRESS

FIRST SESSION

HEARING HELD IN WASHINGTON, DC, MARCH 21, 2007

Serial No. 110–11

Printed for the use of the Committee on Education and Labor

Available on the Internet:
http://www.gpoaccess.gov/congress/house/education/index.html

U.S. GOVERNMENT PRINTING OFFICE

34–015 PDF WASHINGTON : 2007

For sale by the Superintendent of Documents, U.S. Government Printing Office
Internet: bookstore.gpo.gov Phone: toll free (866) 512–1800; DC area (202) 512–1800
Fax: (202) 512–2104 Mail: Stop IDCC, Washington, DC 20402–0001

COMMITTEE ON EDUCATION AND LABOR

GEORGE MILLER, California, *Chairman*

Dale E. Kildee, Michigan, *Vice Chairman*
Donald M. Payne, New Jersey
Robert E. Andrews, New Jersey
Robert C. "Bobby" Scott, Virginia
Lynn C. Woolsey, California
Rubén Hinojosa, Texas
Carolyn McCarthy, New York
John F. Tierney, Massachusetts
Dennis J. Kucinich, Ohio
David Wu, Oregon
Rush D. Holt, New Jersey
Susan A. Davis, California
Danny K. Davis, Illinois
Raúl M. Grijalva, Arizona
Timothy H. Bishop, New York
Linda T. Sánchez, California
John P. Sarbanes, Maryland
Joe Sestak, Pennsylvania
David Loebsack, Iowa
Mazie Hirono, Hawaii
Jason Altmire, Pennsylvania
John A. Yarmuth, Kentucky
Phil Hare, Illinois
Yvette D. Clarke, New York
Joe Courtney, Connecticut
Carol Shea-Porter, New Hampshire

Howard P. "Buck" McKeon, California, *Ranking Minority Member*
Thomas E. Petri, Wisconsin
Peter Hoekstra, Michigan
Michael N. Castle, Delaware
Mark E. Souder, Indiana
Vernon J. Ehlers, Michigan
Judy Biggert, Illinois
Todd Russell Platts, Pennsylvania
Ric Keller, Florida
Joe Wilson, South Carolina
John Kline, Minnesota
Bob Inglis, South Carolina
Cathy McMorris Rodgers, Washington
Kenny Marchant, Texas
Tom Price, Georgia
Luis G. Fortuño, Puerto Rico
Charles W. Boustany, Jr., Louisiana
Virginia Foxx, North Carolina
John R. "Randy" Kuhl, Jr., New York
Rob Bishop, Utah
David Davis, Tennessee
Timothy Walberg, Michigan

Mark Zuckerman, *Staff Director*
Vic Klatt, *Minority Staff Director*

C O N T E N T S

ESEA REAUTHORIZATION: OPTIONS FOR IMPROVING NCLB'S MEASURES OF PROGRESS

Wednesday, March 21, 2007
U.S. House of Representatives
Committee on Education and Labor
Washington, DC

The committee met, pursuant to call, at 10:28 a.m., in room 2175, Rayburn House Office Building, Hon. George Miller [chairman of the committee] presiding.

Present: Representatives Miller, Kildee, Payne, Hinojosa, Tierney, Kucinich, Wu, Holt, Davis of California, Sarbanes, Sestak, Loebsack, Hirono, Yarmuth, Hare, Courtney, Shea-Porter, McKeon, Petri, Castle, Souder, Ehlers, Platts, Keller, Fortuno, Boustany, Kuhl, and Heller.

Staff present: Aaron Albright, Press Secretary; Tylease Alli, Hearing Clerk; Alice Cain, Senior Education Policy Advisor (K-12); Fran-Victoria Cox, Documents Clerk; Adrienne Dunbar, Legislative Fellow, Education; Amy Elverum, Legislative Fellow, Education; Denise Forte, Director of Education Policy; Gabriella Gomez, Senior Education Policy Advisor (Higher Education); Lloyd Horwich, Policy Advisor for Subcommittee on Early Childhood, Elementary and Secretary Education; Lamont Ivey, Staff Assistant, Education; Brian Kennedy, General Counsel; Ann-Frances Lambert, Administrative Assistant to Director of Education Policy; Ricardo Martinez, Policy Advisor for Subcommittee on Higher Education, Lifelong Learning and Competitiveness; Stephanie Moore, General Counsel; Jill Morningstar, Education Policy Advisor; Joe Novotny, Chief Clerk; Lisette Partelow, Staff Assistant, Education; Rachel Racusen, Deputy Communications Director; Theda Zawaiza, Senior Disability Policy Advisor; Mark Zuckerman, Staff Director; James Bergeron, Counselor to the Chairman; Robert Borden, General Counsel; Kathryn Bruns, Legislative Assistant; Steve Forde, Communications Director; Jessica Gross, Deputy Press Secretary; Taylor Hansen, Legislative Assistant; Chad Miller, Professional Staff; Susan Ross, Director of Education and Human Resources Policy; and Linda Stevens, Chief Clerk/Assistant to the General Counsel.

Chairman MILLER [presiding]. Good morning. The Committee on Education and Labor will come to order.

Today's hearing will shed light on one of the most important decisions we face in reviewing the No Child Left Behind law: whether or not to reform the current definition of adequate yearly progress.

I can think of no question more central to the reauthorization and goals of the law.

As one of the original authors of No Child Left Behind, I am often asked how I would like to see the law changed. The short answer is that I would like to see us be responsive to legitimate concerns while maintaining the core values of the law, providing an equal opportunity and an excellent education to every child, regardless of their race, their family income or disability.

I recognize that there are some legitimate concerns with the current accountability system. And today we have the opportunity to focus on two concerns that have been central to this discussion on reauthorization: One, will a growth model system offer real accountability for student achievement? And, two, are there other credible and reliable academic indicators in addition to standardized tests that can offer an accurate picture of student achievement?

With the system that we have currently, what is commonly known as the status model, we know there are some schools where students are making real progress and yet these schools are still not making AYP. Under the current system, a gain or loss in the percentage of students who are proficient could be a result of factors largely outside the school.

At the joint hearings we held in this room last week with members of the House and Senate Education Committees, every organization who testified proposed growth models as the solution to these challenges. Today we will have the opportunity to examine whether growth models are the answer schools and states are seeking.

The second focus of today's hearing has also generated much debate. And that is the concern that a single standardized test is too blunt an instrument to fairly and effectively measure school progress. We have heard from many in the civil rights, education and research communities who acknowledge that using one standardized test to compare students against a single set of high standards is essential to closing the achievement gap.

They have also expressed valid concerns that that single test may not be able to tell us all we need to know about what students and schools can do. Having the most accurate information on student progress is critical to closing the achievement gap. And looking at other evidence in addition to state tests may be the way to obtain a more complete view of a child's true progress.

Further, including indicators such as graduation rates and advanced course taking may incentivize progress in closing the debilitating achievement gaps in those critical areas. Today we will hear from leading experts and practitioners on these two complex accountability issues: growth models and multiple indicators.

I look forward to their testimony and ask them to keep in mind three questions as we look for their help in these areas.

First, are growth models and multiple indicators of performance consistent with No Child Left Behind's goal of ensuring that all children can read and do math at grade level by 2014?

Second, do states have the capacity they need to ensure that information gathered to determine whether a school or district has made adequate progress is both valid and reliable?

Third, do these approaches appropriately credit improving schools, or do they overstate academic progress? In other words, are they a step forward in offering a fairer, more reliable means of accountability, or are they a step backward, simply another loophole that hinders accountability?

Our collective goal in reauthorizing No Child Left Behind should be to look to those changes that improve the integrity of the act and move us forward toward the stated goal of the act, to provide opportunity and an excellent education to every child.

I want to thank the witnesses in advance for their testimony.

And I would like now to yield to the senior Republican on the committee, Mr. McKeon, for his opening statement.

Mr. MCKEON. Thank you, Mr. Chairman, for convening this hearing as part of the series of hearings on No Child Left Behind we launched a week ago.

Though last week's discussion provided a broad overview of our reauthorization effort and gathered input from both the House and the Senate, I believe today's hearing and the others to follow will serve an even more important purpose as they delve into the real challenges at the heart of NCLB.

Today we begin with an examination of options for improving NCLB's measure of progress. And I thank our panel of witnesses for joining us for this examination.

Adequate yearly progress is a benchmark that makes NCLB different from other education laws that came before it. It is the measure that tells all of us legislators, parents, teachers, administrators, and taxpayers exactly how a school is doing in educating students from one grade level to the next. And for that reason, it is vital that the concept remains in place.

However, as we approach this year's reauthorization, it is important that we are open minded to tweaks in the law that could make it more practical while ensuring that the underlying principle of accountability remains consistent. And that is where growth models enter into this discussion.

Under current No Child Left Behind guidelines, school districts use a status model to compare the performance of students in a specific grade against the performance of the students of that same grade during the previous year. Some have raised concerns about the reliability of the status model. They argue that a model which compares the achievement of the same students over time within a growth model may be more appropriate and act as a more accurate measure of adequate yearly progress.

As we review the Department of Education's growth model pilot programs as well as last year's Government Accountability Office report on the implementation of growth models to determine if schools in certain states were making adequate yearly progress under No Child Left Behind, I believe that growth models can play an important role in this reauthorization. However, these growth models must be well-designed. They must be rigorous. And they must meet a number of criteria that are consistent and central to NCLB.

For example, they must include the requirements that all students reach proficiency, that the gaps between groups of students continue to close, and the growth model is tracked as part of a

state data system and that a state's assessment system must produce comparable results from grade to grade and year to year.

With that being said, members of this committee know as well as anyone that the reliability and utility of growth models is the focus of an ongoing debate. So I think we all can comfortably say today that we are not necessarily here to wholeheartedly embrace the concept nor dismiss it out of hand. Instead, we are simply here to listen and to learn. I am looking forward to this hearing and the additional hearings we will be having in this series.

And again, I thank the witnesses and look forward to their testimony.

Chairman MILLER. Thank you.

With that, we will begin with the witnesses.

Our first witness will be Allan Olson, who is the co-founder and chief academic officer of the Northwest Evaluation Association. Northwest Evaluation Association is a non-profit organization that provides research, support and technical assistance to 2,400 partnering school districts and education agencies throughout the United States. Dr. Olson has led the Northwest Evaluation Association in its efforts to build the largest nationwide database in longitudinal student test results.

Valerie Woodruff is the secretary of education from the Delaware Department—excuse me. I think our colleague wanted to introduce——

Mr. CASTLE. Thank you, Mr. Chairman.

It is a great pleasure for me to introduce Delaware's secretary of education, Valerie Woodruff. Val has been secretary since July of 1999, prior to which she served as the associate secretary for curriculum and instructional improvement for Delaware. Her career is rooted in education. And she has been a teacher, counselor, assistant principal and principal in high schools in both Maryland and Delaware.

As secretary, Val has led the implementation of Delaware's accountability system as well as implementation of No Child Left Behind. I appreciate Val's commitment to raising student achievement, the importance of high-quality teachers and school leaders and the belief that all children deserve an excellent educational experience.

Val is the Delaware representative on the Southern Regional Education Board, serves on the executive committee of SREB and is the first K through 12 educator to serve as vice chair. She has also served on the Board of the Council of Chief State School Officers and was the president of the Chief State School Officers from November of 2005 through November of 2006.

We are lucky to have her in Delaware. And don't try to take her away.

I yield back.

Chairman MILLER. Thank you.

Welcome, Secretary Woodruff.

Dr. Chrys Dougherty is the associate director of Research National Center for Educational Accountability. Dr. Dougherty is the director of this center and has authored the "Parents Guide to Asking the Right Questions about School" and has written extensively

on the value of longitudinal data and the 10 essential elements of statewide student information systems.

He has been an elementary school science teacher in Oakland, California, is a professor of statistics, econometrics—ergonomics is what we fight over in the labor side of this committee.

Mr. DOUGHERTY. Yes, econometrics.

Chairman MILLER. Econometrics. Yes. You are the guys? They are always quoting you guys about this and that. Okay—and education policy at the LBJ School of Public Affairs.

Peter McWalters is commissioner of the Rhode Island Department of Elementary and Secondary Education. Prior to becoming Rhode Island's commissioner, he served over 20 years in a variety of educational leadership and teaching positions, including the superintendent of schools in the city school district of Rochester, New York.

Dr. Harold Doran is the senior research scientist, American Institutes for Research, where he supports the development of state testing and accountability systems as an applied statistician and psychometrician. And he is currently a member of Secretary Spellings' peer review panel for state growth models. He has been an elementary school principal and a classroom teacher.

Welcome to all of you, and thank you for your contributions this morning.

Mr. Olson, we are going to begin with you.

There will be a light in front of you. The green lights will go on when you start your testimony. There will be a yellow light that suggests you should start wrapping up in the next minute or so, and then a red light when your time has run out.

But we will obviously allow you to finish a thought and a sentence and maybe even a paragraph. There you go.

STATEMENT OF ALLAN L. OLSON, CO–FOUNDER AND CHIEF ACADEMIC OFFICER, NORTHWEST EVALUATION ASSOCIATION

Mr. OLSON. If it is brief.

Chairman Miller, Ranking Minority Member McKeon and members of the committee, I appreciate the opportunity to testify before you.

Again, my name is Allan Olson. I am co-founder of an organization called the Northwest Evaluation Association. The Northwest Evaluation Association is a not-for-profit organization. We provide testing services to school districts around the nation and also have a very strong research staff. So we do research in the field also.

We are currently providing very accurate measures, assessments, and growth measures for approximately 3 million children multiple times a year in 49 states. After 30 years of experience in research, it is clear that NCLB could be strengthened and more effective if states were allowed to and encouraged to implement measures of student achievement that were accurate enough to actually measure growth, actually measure growth of the individual students. Okay?

So I am talking about student level and actually designed specifically for determining change over time at the child level. An accurate growth measure provides the best evidence of a school's effectiveness. It also improves the assessment data in ways that help

students, teachers, parents, and others focus learning and focus their efforts to improve learning over time.

In other words, a very good accurate measure and a growth measure will inform many people within the education community in manners that allow them to change their behaviors to become increasingly effective. So a good growth measure is not only probably the best accountability measure, it is also the best possible way to improve our capacity to improve learning.

Today's computerized adaptive tests represent the most common approach to meet these requirements. However, states could develop other methodologies.

An accurate measure of each student's achievement is reported on a cross-grade vertical measurement scale provides the school and the state information about whether a student is proficient, in other words, meets all the requirements inside No Child Left Behind related to status capacity and information about how far a student is below or above that standard, not just information that the child is below or above, but actually how far below or above, which also gives us a chance to establish growth targets at the child level, growth targets that would lead toward proficiency and/or growth targets that would help children or focus on children who are well above the standard at the time of the measure.

Allowing states to accurately measure growth of each child strengthens all the foundation pieces of No Child Left Behind while providing educators evidence that will inform improvement of instruction and learning. So what we would be asking for is states be allowed to have a system that increases the quality and accuracy of information to inform the process's improvement while putting in place an accountability measure as required by No Child Left Behind.

As I mentioned before, growth measure is the best measure of whether a school, a program, a district is being effective in meeting the needs of its children. A growth measure that is accurate enough to measure growth in an individual child also helps a district know whether they are being effective with children of differing characteristics, whether that is ethnicity, whether it is gender, whether it is starting place on a scale. It gives the school district information about how effective they are with those children.

Accuracy is the center piece of a good growth measure. The test requirements today that are in place, the tests the states have in place today are quite accurate for children who happen to be near the proficient line or happen to be in the middle of a distribution. But by the nature of the design requirements for tests, the tests that are in place will not be as accurate for low-achieving children or will not be as accurate for high-achieving children.

And if your measure isn't as accurate for low-or high-achieving children, it will not be a best growth measure for those children and will not provide the kind of information that will lead to constant improvement focussed on those children. A measure of that nature also will not be very accurate for purposes of diagnostic reporting, which is one of the requirements of No Child Left Behind. But states probably are falling short of the intent of that particular provision in the law. A good, accurate measure, a good, accurate growth measure would allow states to respond in that manner.

I think in order to have a very good measure, it will be important to remove the real tight constraints right now that are in place, either intentionally or by the nature of the way the law is being implemented, remove the constraints for a very tight alignment to just grade level content standards with the measure. Many children are functioning well below those content standards. And we need to measure those children well.

The law calls for challenging all children. To challenge all children, we must have a measure that is accurate for all children and be able to set growth targets that are appropriate for those children.

Thank you very much.

[The statement of Mr. Olson follows:]

Prepared Statement of Allan Olson, Co-Founder and Chief Academic Officer, Northwest Evaluation Association

The Northwest Evaluation Association (NWEA) is a not-for-profit organization which partners with over 2,500 school districts to promote student learning provides, precise and consistent growth assessment testing services for over 3 million children in 49 states. For over 30 years, we have been providing assessments in key subjects in grades 2-12, as well as detailed reports on student learning, and offering training to help educators use data to improve practice. Our tests are given multiple times per year in paper-and-pencil and computer-adaptive formats and give educators, parents, students, and policymakers a clear and comprehensive look at how much academic growth individual students are making over time. This kind of data has been of great value to our partner districts and has resulted in increases in the number of children tested at a rate of over 50 per cent per year. NWEA's mission—"partnering to help all kids learn"—also has lead us to research educational policy and practice based on the extensive data in our database and our experience with thousands of teachers and schools.

In the course of this research and working with our 2,500 partner districts, it has become clear to us that in order to help students learn more, we have to provide teachers with the information that they need to be able to identify student strengths and deficiencies and to better understand how far each child is from achieving proficiency. This means that we have to measure accurately each student's current achievement level to understand what a student knows and needs to know next, and to track each student's growth over time to be sure that young people are moving at a rate of growth that will help them become proficient. We have to provide this information to the teacher as quickly as possible, in a form that enables the teacher to make the best instructional decisions for the students.

The aspect of this approach that is germane today is the measurement and use of student growth information. What we mean by growth measurement is using assessment to "follow the child" in order to find the actual achievement level of the child and then to measure it over time.

In this area, our organization has reached three conclusions, as follows:

1. We will gain a much more complete and useful picture of the performance of our schools if we include the growth of individual students in our accountability systems.

2. Students must have growth targets that challenge them and that lead them to the state's definition of proficiency in a set of skills that will make them productive members of society when they graduate from high school.

3. Teachers, principals, students, and parents must all have a clear understanding of the amount of achievement growth that the student must make each year to enable them to participate in the student's growth.

Why is measuring individual achievement growth important?

As NCLB has been implemented, it has become increasingly obvious that the way student achievement is measured currently does not begin to tell us whether the school is doing a good job or a poor job teaching the students that come through its doors. While there are many reasons for this, the issue can be seen very clearly as follows:

Schools "A" and "B" have the same percentage of students identified as "proficient." Students in school B grew, on average, twice as much as students in school A to achieve their proficiency. Which school is doing a better job?

We believe that the answer is the school that is achieving greater rate of progress in moving students towards proficiency. Promoting the growth of individual students from one year to the next is the hallmark of a successful school. This is especially true for students who are below proficiency levels for a given grade and need to grow faster in order to catch up. Providing teachers a measure of how much the student must grow to get where the students needs to be also gives that teacher a useful tool for addressing the learning needs of each individual student.

Students come to school with different preparation, motivation, and support resources. It is the job of every school to help every student move forward regardless of his or her current achievement level. For students with low achievement levels, the school needs to accelerate growth, to help these students reach levels that will allow them to compete when they graduate from school. For students with high achievement levels, the school needs to keep them growing to keep them engaged and to allow them to reach their full potential.

Research (Kingsbury and McCall, 2006) has clearly indicated that schools vary greatly in the amount of growth that they cause in student achievement. It is equally clear that student growth differs by grade and demographic group within a school. Without information about student growth, we cannot tell the full story of a school, and we shouldn't try to judge whether the school is doing a good job or not.

Can we measure achievement growth of individual students?

It is clear that two components are needed to measure the achievement growth of individual students. The first requirement is the ability to measure students accurately to gain a deep understanding of where their learning is. Current tests provide very little information about students who are high performers and are well beyond their grade level or low performers who are well behind grade level. To be able to measure achievement for these students requires a measurement scale that goes beyond grade-level testing and identifies what students know across the many strands of knowledge that a student needs to know to be identified as a proficient.

Let me illustrate the point. Consider, for example, a twelve year old child (grade 6) performing two grade levels below his age level (grade 4). If that child achieves a year and a half of growth for each of the next two years, he will be in grade 8 and perform much like a 7th grade student. That is a huge success. However, if we only measure the "status" of the child as to his age level, and not the growth, we will conclude that the child is a failure and the school is failing him even though he will have caught up a whole grade level. Further, we won't be able to inform the teacher, the parents, or the child where the student is truly performing so that they can craft a plan to reach proficiency.

The tools are available to provide this kind of detailed information. Growth measures have been in use for several decades. Computerized adaptive testing (CAT: Weiss, 1982) was developed by researchers with funding from the federal government in order to provide a way to measure large, diverse groups of individuals efficiently and accurately. An adaptive test allows us to measure the performance of high-achieving and low-achieving students as accurately as we measure the students in the middle of the distribution. Since its development, adaptive testing has been used for a host of high-stakes and low-stakes applications, from individuals entering the armed services to individuals trying to be certified in high-tech specialties. NWEA alone has administered over 60,000,000 adaptive tests to students.

NWEA urges Congress to allow states and school districts to measure student growth as part of the accountability requirements under No Child Left Behind. We believe the great advantages such an approach provides will be sufficient motive to states to adopt this option as they consider how best to serve their children.

It is important to stress that we are not proposing to abandon information about whether a child is operating at grade level. Rather, we want to allow states to go further. As illustrated in the slides that accompany this testimony, we can be far more effective in helping children achieve greater growth, so they can move to proficiency and beyond, if we more accurately know where they are performing and we can measure their performance growth.

What Measuring Growth Can Do

One of the critical challenges confronting NCLB is ensuring that accountability is linked to approaches that actually are useful in helping schools and teachers help students reach proficiency.

If we know where a student stands, and how much they must grow before they graduate, we should be able to marshal our resources to make sure that the needed growth occurs.

If we know how much growth is typical for a student who starts the year with a certain level of achievement, we should be able to immediately set goals for the student that represent good growth, great growth, and incredible growth.

If we know the growth goals for a student, we should be able to tell the teacher exactly what the student needs to learn by the end of the year to meet the growth goals.

If we know the growth goals for each student that a teacher is working with, we should be able to guide that teacher so that he or she can design and redesign the instructional approach she will take with her students.

And if we accomplish these things, the accountability is aligned with how students learn and what schools need to do.

After all, the central issue is how we help the current generation of students meet our expectations. Measuring growth of each child gives us information that we can use to improve the growth of all of our students. At the same time, information about growth at the class and school level helps us describe our schools and their efficiency in ways that are far more useful to schools, teachers, parents and kids than what we learn by confining ourselves to the simple status question of current grade level.

Finally, for our students who aren't growing to meet their growth goals, our response needs to be centered on the needs of those students. We need to reorganize to help the students.

In conclusion, our request is a simple one: make it clear in the law that states are permitted, or even encouraged, to do more than just measure status. They can, and should, also measure growth as part of that same process.

Thank you for the opportunity to share our experience and data with you this morning.

Improving NCLB Accountability

Current Law: NCLB requires states to develop a measure of annual yearly progress (AYP) in order to hold districts and schools accountable. It stipulates that by the 2005-06 school year the states must have in place an assessment system for all students, as well as various subgroups, that annually tests student performance in reading/language arts and mathematics in grades 3-8, and for a single test in grades 10-12. By the 2007-08 school year, states are also required to assess every student in science, at least once in each of the following grade spans: 3-5, 6-9, and 10-12.

NCLB also allows states and localities to include other measures of student academic progress but these measures may not be used in place of the assessments described above for purposes of establishing AYP.

The Problem: Currently under NCLB, schools are evaluated for their progress in improving student performance by comparing successive groups of students rather than tracking the same group of students over time. In other words, to meet AYP, schools must show that each grade level (e.g. third graders) has improved over the previous year, not that each student or the same group of students (e.g. third graders that are now fourth graders) has progressed. Therefore, these yearly comparisons do not track the performance of the same students.

This approach to assessment does not provide the information we need to accurately measure what individual students know and what educators need to know to address their learning deficiencies and support their achievement growth.

In addition, since the focus of NCLB is on measuring proficiency rather than annual learning progress, schools that have improved substantially but have not yet reached proficiency targets are rated the same way as schools that have no improvement. Achieving learning gains provides no credit to these schools.

The Solution: In addition to the annual testing by grade and by subject currently required, states should be allowed to meet their NCLB annual yearly progress assessment requirements by measuring the performance growth of every student.

NCLB recognizes the critical role that timely, accessible, and accurate information about student academic performance plays in informing and motivating educators, policymakers, parents, and the public in finding ways to raise student achievement and close the achievement gap. Giving states the option of measuring student growth to meet AYP assessment requirements would provide a more accurate measure of how students are progressing. By measuring growth over the course of each grade, it would provide educators a clear roadmap for bringing a student to proficiency.

Currently, schools that improved substantially but did not make AYP are viewed the same way as schools that made no improvement. Including a growth measure in assessing school improvement would be fairer. Schools that have made substantial gains in student academic performance would be recognized for those improve-

ments, even if they still do not meet proficiency standards. This change also would allow states to focus their support on those schools that are really struggling.

Questions and Answers

What are the key attributes of a growth model of assessment?

Growth measures provide the kind of information about what students know and do not know in key strands of knowledge within subject areas that helps teachers identify and focus on student strengths and deficiencies and determine what needs to be taught next. Using growth models, educators and young people can identify desired semester-by-semester targets for student achievement that, if met, will ensure that young people are making progress toward mastering content and attaining proficiency. With this information, proficiency targets are not some abstract, far-away goal but clear benchmarks for students and teachers to reach that help ensure that students achieve proficiency over time.

Measuring growth requires testing students against a common scale. This means that student achievement is measured to determine where a student fits across the entire continuum of learning in a particular subject area rather than on a grade-specific scale. The growth measure is actually a measure of growth toward proficiency, which is not tied to grade level but to mastery of content. Tests used by states today that measure what a student needs to knows within a particular grade level provide very good information about students performing in the middle range of performance (where state cut scores for accountability are pegged). But these tests do not ask enough questions to paint a useful portrait of what is happening with high-achieving and low-achieving young people who typically perform at the extremes or outside their grade levels. For example, state grade-level tests provide little information when a sixth-grade student is performing at the fourth-grade level or about a fourth-grade student who is performing at the fifth- or sixth-grade levels.

If a state chooses to measure student performance growth from year to year instead of progress towards meeting fixed performance targets, won't the gaps between low- and high-performing students just be continued?

Not necessarily. If states set growth targets on the road to proficiency then states, districts, and schools will continue to have markers to meet to ensure that all students graduate from high school with the knowledge and skills they need for productive and success lives.

Is it realistic to assume that low-performing students can grow at a faster rate than higher-performing students to meet those targets?

Currently, NCLB requires states, districts and schools to meet fixed performance targets by grade and by subject for all children. The only way to meet the intended purpose of NCLB—to close achievement gaps—is to identify those gaps and develop strategies for addressing them. By providing schools and teachers information on how a student is progressing within the school year and between school years is more likely to impact teaching and learning and, therefore, accelerates improvements in student achievement.

Using growth measures also addresses another key problem with the current law. Currently, state targets for AYP are set all over the map. While a few states have set high performance targets early on, many are waiting until several years from now to establish higher targets for achievement that are closer to desired proficiencies. This delay means that in several years, schools that have been judged as meeting AYP will suddenly be far off from state targets. Growth measures provide a way of setting steady and achievable targets that are based on what can truly be expected of young people.

How are student growth measures different than the currently used value-added testing, also called a "growth model"?

The U.S. Department of Education is supporting pilot "growth model" accountability plans in school districts in Arkansas, Delaware, Florida, North Carolina, and Tennessee. It has been mandated for use by all school districts in Pennsylvania and Ohio and several hundred school districts in 21 states. New legislation in Arkansas and Minnesota calls for implementing a form of value-added measurement, and the School Boards Associations in Iowa and New York are currently piloting a value-added program. Dallas and Seattle are the most prominent urban districts that use the value-added approach. In some states, such as Tennessee, this value-added model (VAM) is the bedrock of the accountability system, and the results are used to judge the quality of schools and the effectiveness of individual teachers.

Value-added models of assessment, however, are an analytic methodology applied to NCLB test results. It is a method of statistical analysis, rather than a particular test, used to analyze longitudinal test data in order to isolate factors affecting a student's growth over time. It provides educators general information about which students have benefited most and least and about instructional impact—how effective

it has been in providing students with a year's worth of growth from where they began the year. Through this information, teachers, principals, district administrators, and school board leaders can learn whether high achievers, middle achievers, or low-achievers are making the most progress, and what can be done to raise the performance of each group. Impact data can determine whether and the extent to which schools and classroom teachers are effective in raising performance.

The currently used value-added models, however, do not provide the kind of rich multiple-times per year diagnostic information about the key strands of knowledge within subject areas that each student needs to master to move to the next level of performance. It also does not tell why a particular teacher is effective or not effective. And the value-added analysis is applied to tests that are not particularly accurate for students who are high achievers and low achievers, thus blunting its value as even a broad analytic tool.

Won't a growth model require a sophisticated data system that will substantially add to state and district costs?

States will be given the flexibility to continue with the current assessment models or to substitute or add a growth measure of progress towards measuring AYP.

Our experience suggests that using a growth measure of progress could cost less, not more, than the current NCLB testing requirement. In Idaho, for instance, the cost is $13.00 per students to test students in grades 2-10, four times a year, including training and reporting costs. This is less than most states are spending on once-a-year testing under NCLB requirements.

Isn't testing itself the problem, imposing unnecessary burdens on school districts and leading teachers to teach to the test? Shouldn't we just eliminate the testing requirements from the law?

If the nation is serious about accountability in education and about making sure that tax dollars invested in education result in a student population that is prepared for work and postsecondary education, we should not back away from the concept of testing. The issue is not whether or not to test but what kind of testing will yield the kind of information that actually helps teachers help students. Expansion in the use of growth measures rather than one-shot grade-level tests can help educators, policymakers, and parents determine whether schools and students are actually making required progress toward proficiency. They also will tell educators, school board members, parents, and students what areas of learning they need to be working on to make desired growth targets.

Since more than 2,500 school districts use out of grade-level testing currently, why does the law need to be changed? Can't districts simply do what you're proposing under current law?

Yes, any district can use whatever test it wants to measure student learning. However, the law makes specific reference to use of grade-level tests without referring to growth measures to fulfill the assessment and accountability requirements of NCLB. The 2,500 school districts that use growth measures to determine the performance growth of children are pioneers that have demonstrated the value of this kind of assessment to provide comprehensive information about individual student achievement in key subject areas to help further accelerate achievement gains. There are over 12,000 other public school districts that might include testing that tracks the performance of students over time if the law explicitly recognized this kind of testing as an alternative in determining whether schools, districts, and states are in compliance with the law.

Do other companies also offer this sort of testing, or are you simply trying to change NCLB to benefit NWEA?

Many other testing organizations—such as the Educational Testing Service and Scantron—already use testing methodologies that can pinpoint individual student achievement against a common scale and provide immediate feedback. This type of testing was first introduced by the U.S. military in the 1970s. The computer-adaptive testing used by NWEA, for example, is basically the same methodology ETS uses in its Graduate Record Exam and GMAT tests.

Encouraging states to use computer-adaptive methodologies and growth measures that can given by computer or paper-and-pencil tests might actually hurt NWEA by providing much larger companies greater incentives to develop growth measures and enter this market. But we believe that it is the right thing to do and is not simply a matter of which companies have the biggest market share, but whether we have the kinds of tests that will help more schools bring more students to proficiency.

Northwest Evaluation Association

The Northwest Evaluation Association (NWEA) is a national nonprofit organization based in Portland, Oregon, that partners with school districts and education

agencies nationwide to promote academic student growth and school improvement. NWEA provides computer adaptive and paper-and-pencil assessments in mathematics, language arts, and science in grades 2-12 as well as training and comprehensive reporting tools that enable educators to measure and promote individual student and school academic growth. Their products and tools are provided at a price districts can afford, and any profit is reinvested in product development and technical assistance.

Three decades of experience nationwide. Over the past 30 years, the company has tested more than 25 million young people; it currently is helping to assess more than 4 million students a year in more than 2,400 school districts in 49 states. Its presence is particularly strong in Illinois, Indiana, Minnesota, New Hampshire, and South Carolina, where it tests the vast majority of students in the state.

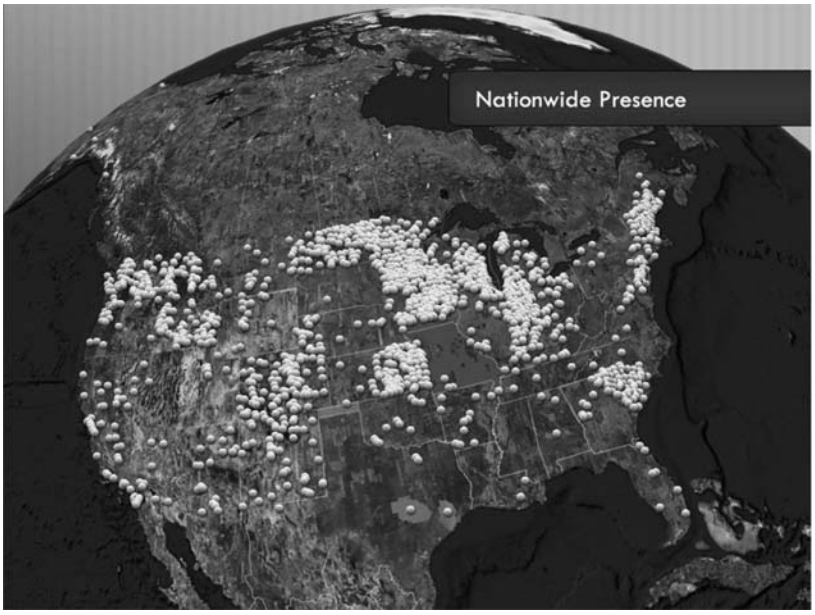
Growing demand for student growth data to support NCLB. NWEA has grown by 50 percent a year in recent years to meet the demand of school districts for formative assessments that track the growth of individual students over time and offer immediate feedback to district leaders, teachers, students, parents, and school board members.

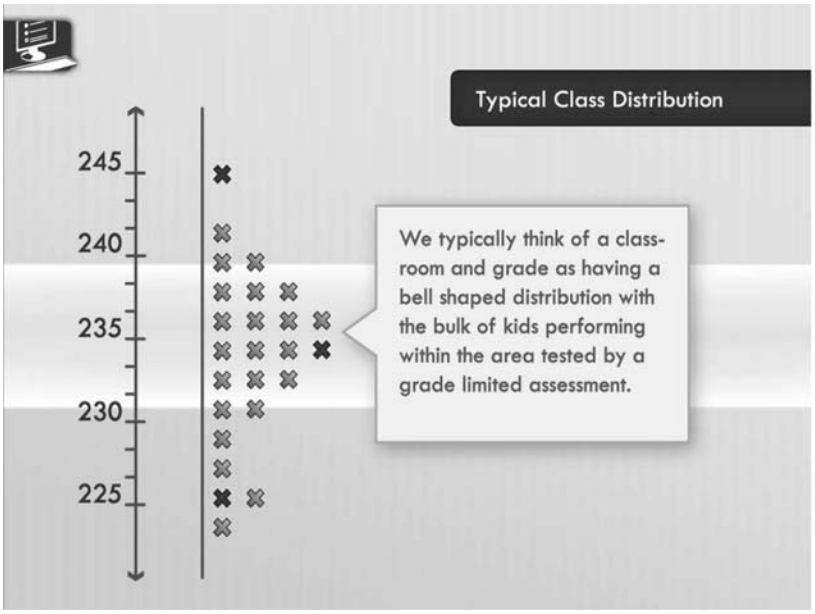
An immediate and vital source of information for teachers. The value of the assessments to schools is considerable, in part because students and teachers receive immediate results which allow them to better understand and develop strategies to offset student learning deficiencies. The assessments evaluate student achievement across content standards, and results help identify problem areas in content knowledge, skills, and concepts that need addressing to best maximize achievement. Because it is a growth measure, teachers use the data to determine if students are making equal to, or normal, growth. The test also offers schools valuable information about the most effective teachers, student groupings, or the need for alternative ways to focus instruction.

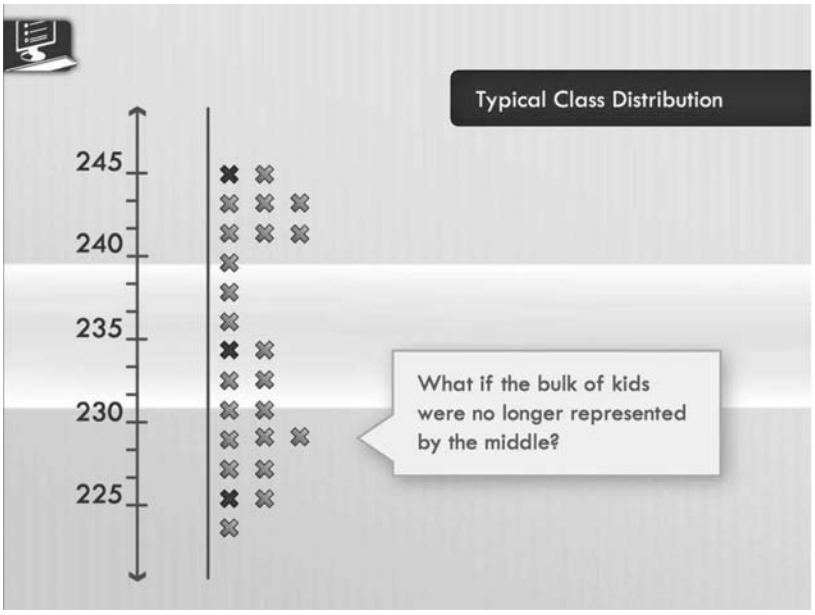
An accountability tool for NCLB. In addition, schools and district leaders can compare scores with the growth targets for a particular year and see whether students are on target for meeting proficiency levels required to achieve the goals of NCLB. Results can be disaggregated by NCLB subgroups to give periodic indicators about how well a school is doing in serving diverse populations.

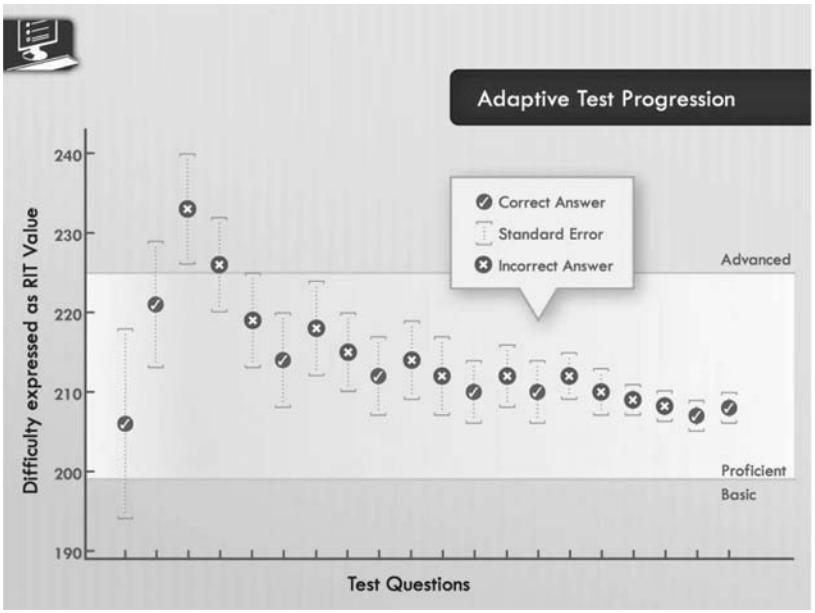
A unique resource for finding proven answers to some of our most challenging educational issues. All the student growth data gathered by NWEA is aggregated into our Growth Research Database, the largest nationwide repository of student test results which is used by states, national organizations, and prominent national researchers to assess the impacts of policy and practice on student achievement growth.

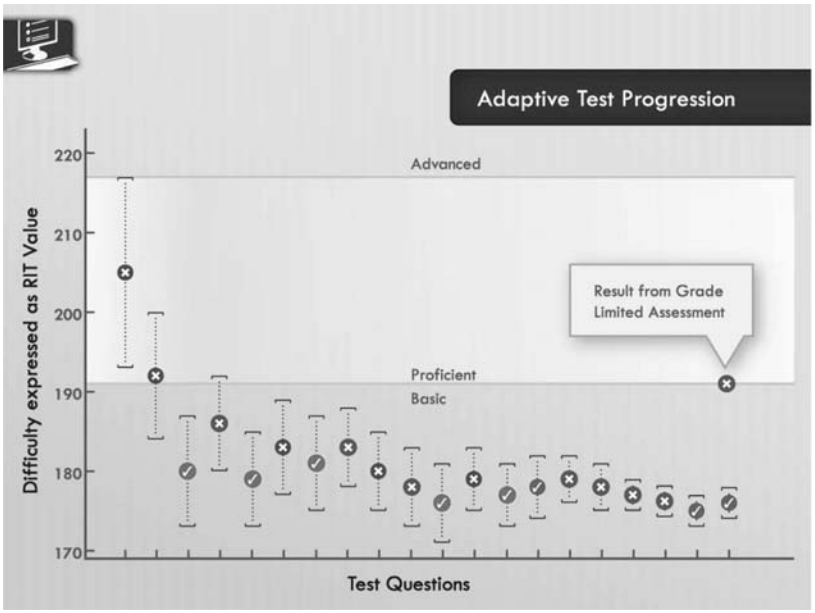
For more information, go to *http://www.nwea.org*.

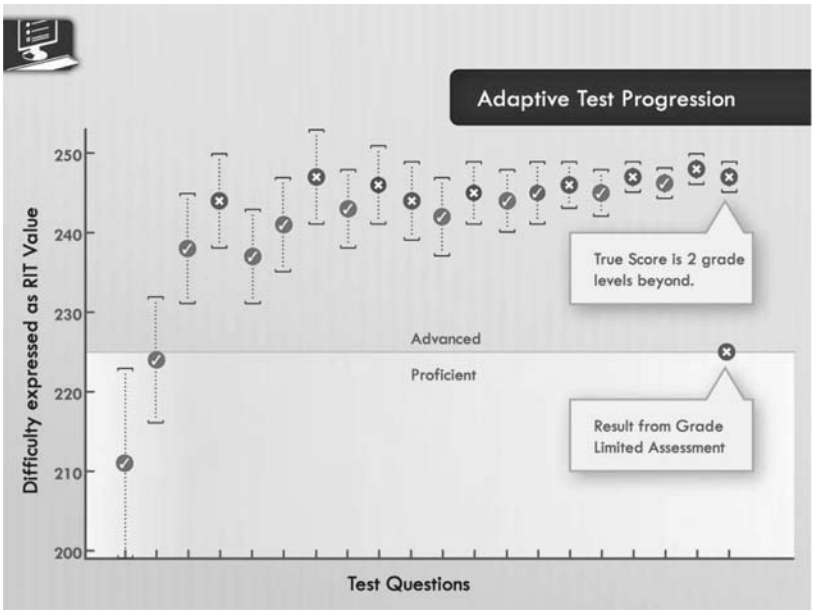

Nationwide Presence

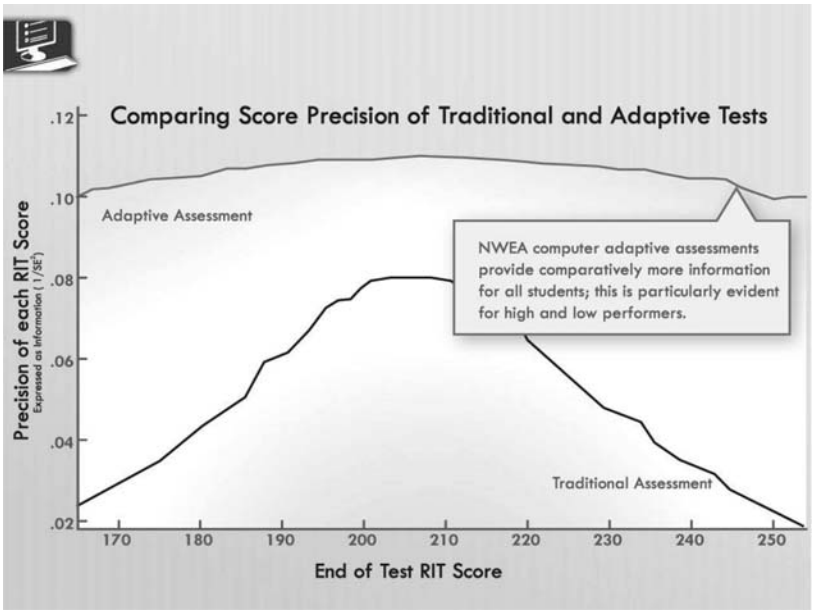

Typical Class Distribution
245
240
235
230
225
We typically think of a class-room and grade as having a bell shaped distribution with the bulk of kids performing within the area tested by a grade limited assessment.


Typical Class Distribution
245
240
235
230
225
What if the bulk of kids were no longer represented by the middle?


Adaptive Test Progression
Difficulty expressed as RIT Value
240
230
220
210
200
190
Correct Answer
Standard Error
Incorrect Answer
Advanced
Proficient
Basic
Test Questions


Adaptive Test Progression
Advanced
Proficient
Basic
Result from Grade Limited Assessment
Difficulty expressed as RIT Value
220
210
200
190
180
170
Test Questions


Adaptive Test Progression
True Score is 2 grade levels beyond.
Advanced
Proficient
Result from Grade Limited Assessment
Difficulty expressed as RIT Value
250
240
230
220
210
200
Test Questions



STATEMENT OF VALERIE WOODRUFF, SECRETARY OF EDUCATION, STATE OF DELAWARE

Ms. WOODRUFF. Good morning.

Chairman MILLER. I am going to have to turn my microphone on, and you are going to have to pull yours closer to you.

Ms. WOODRUFF. Okay. It wasn't on. Okay?

Chairman Miller, Ranking Member McKeon, and members of the committee, thank you for this opportunity to testify today about the implementation of growth models and accountability systems.

I am proud to say that Delaware was among several states that had implemented a school and district accountability system to measure our progress and standards-based reform prior to the passage of No Child Left Behind.

We began assessing English language, arts, and mathematics in 1998. And based on early information about the goals of No Child Left Behind, we applauded the initial work of Congress and believed that we could easily meet the requirements of the law.

Our original accountability system included three measures of student performance: status, which is essentially AYP; growth; and the improvement of the lowest performing student. To our schools and to our communities, these measures made sense and had what I refer to as face validity. Simply stated, educators and others understood the value of measuring not only the performance of one cohort of students to another, but also the change in performance of the same cohort of children over time.

And certainly, they saw the value of attending to and measuring the improvement of the lowest performing students and of closing the achievement gap. Delaware was the tenth state to receive approval of our accountability plan in the spring of 2003. Also, we

were among the first states to receive full approval of our standards and assessment system.

Delaware implemented a unique identifier in 1984. And we have worked diligently since that time to link student demographic data with achievement data. And we have reported that for many years.

Given all of these factors, we were anxious to talk with the Department of Education and to convince them that the use of growth models was a natural progression in creating a mature accountability system. When the department allowed states to submit growth models for the 2006 accountability measurement, we felt confident that our proposal would be approved. That did not occur. And we were perplexed at the feedback we received.

None of the questions were related to the model itself. They had to do with other things. It did not seem that the peer reviewers had clear guidance about the criteria, nor did they understand the different models that can be used to measure growth.

We made several changes. And we were approved and will be using the growth model measurement for the 2007 accountability year.

The model that we chose supports our philosophy of continuous improvement for all students. It is easy to understand. It is easy to explain. It provides schools with information that shows which students are making progress toward proficiency, which students are maintaining proficiency, and which students are slipping backwards, which is something we all want to avoid.

It is not enough to measure the average performance of even a small cohort of students. Systems must focus on the performance of individual students and must provide schools with the appropriate incentives to address student needs.

Moving forward, the law should not only encourage the use of a variety of accountability models, not only allow it, but also encourage it. These models should be focused on individual student achievement and build on adequate yearly progress to promote more valid, reliable, and educationally meaningful determinations. States need to be encouraged to innovate and seek new and better ways of continuous student achievement.

Specifically, the Department of Education must establish clear and consistent policies and procedures that enable states to use growth models. It should articulate the foundation elements that must be in place. For example, the state must have a unique student identifier, approved standards and assessment system, and a data system that is able to collect and track individual performance over time.

When states have those elements in place, they should not then have to guess about how their proposals will be judged. Those criteria need to be clear and understandable. They should define what must be contained, and they must select and train peer reviewers so that states can be guaranteed a fair and equitable review of all proposals, regardless of the background or philosophical beliefs of the reviewers. The peer review process must be transparent and iterative and be focused on improving the quality of the accountability system, not limiting their scope and use.

In order for states to pursue stronger, more robust systems of accountability, a partnership of support and technical assistance

must be in place. States need ongoing technical assistance in order to build a strong knowledge base about accountability models. We need to benefit from research about which models are most effective and why. And they need continuing support and development and improving of data systems.

For example, as strong as our data system is today in Delaware, we can benefit from knowledge and support about cutting edge technology. All states are eager to learn more and to improve the quality of education for all of our children.

I appreciate the opportunity. And I will be glad to answer questions. Thank you.

[The statement of Ms. Woodruff follows:]

Prepared Statement of Valerie Woodruff, Secretary of Education, State of Delaware

Chairman Miller, Ranking Member McKeon and members of the committee, thank you for this opportunity to testify today about the implementation of growth models in accountability systems. My name is Valerie Woodruff. I am the Secretary of Education in the state of Delaware. I am the immediate Past President of the Council of Chief State School Officers.

I am proud to say that Delaware was among several states that had implemented a school and district accountability system to measure our progress in standards based reform prior to the passage of No Child Left Behind. We began assessing English language arts and mathematics in 1998. Based on the early information about the goals of NCLB, we applauded the initial work of Congress and believed that we could easily meet the requirements of the law. Our original accountability system included three measures of student performance: status, growth, and improvement of the lowest performing students. To our schools and to our community, these measures made sense and had what I refer to as "face validity." Simply stated, educators and others understood the value of measuring not only the change in performance of one cohort of students to another but also the change in performance of the same cohort of students over time. And certainly, they saw the value of attending to and measuring the improvement of our lowest performing students and of closing the achievement gap.

Delaware was the tenth state to receive approval of our accountability plan in the spring of 2003. Also, we were among the first states to receive full approval of our standards and assessments. Delaware implemented a unique student identifier in 1984 and has worked diligently and deliberately since that time to link student demographic data with achievement data. Given all these factors, we were anxious to engage the Department of Education and to convince them that the use of growth models was a natural progression in creating a mature accountability system.

When the Department allowed states to submit growth model proposals for the 2006 accountability measurement, we felt confident that our proposal would be approved. That did not occur, and we were perplexed at the feedback we received. It did not seem that the peer reviewers had clear guidance about the criteria, nor did they understand the different models that can be used to measure growth. We were required to make several changes in order to receive approval for the 2007 accountability year.

The model that we chose supports our philosophy of continuous improvement for all students. It is easy to explain and understand. It provides schools with information that shows which students are making progress toward proficiency, which students are maintaining proficiency, and which students are slipping backwards. It is not enough to measure the average performance of even a small cohort of students. Systems must focus on the performance of individual students and must provide schools with the appropriate incentives to address student needs.

Moving forward, the law should not only allow but also encourage the use of a variety of accountability models. These models should be focused on individual student achievement and build on adequate yearly progress (AYP) to promote more valid, reliable, and educationally meaningful accountability determinations. States must be encouraged to innovate and to seek new and better ways of supporting continuous student achievement.

Specifically, the Department of Education must establish clear and consistent policies and procedures that enable states to use growth models for accountability. It should articulate the foundation elements that a state needs to have in order to

qualify to use a growth model. For example, a state must have a unique student identifier; approved standards and assessment systems; a data system that is able to collect and track individual student performance over time. When states have those elements in place, they should not have to guess at how their proposals will be judged.

The Department should clearly define what criteria must be contained in a growth model proposal, and they must select and train the peer reviewers so that states can be guaranteed fair and equitable reviews of all proposals regardless of the background or philosophical beliefs of the reviewers. The peer review process must be fully transparent and iterative and be focused on improving the quality of accountability systems, not limiting their scope and use.

In order for states to pursue stronger, more robust systems of accountability, a partnership of support and technical assistance must be in place. States need ongoing technical assistance in order to build a strong knowledge base about accountability models. They need to benefit from research about which models are most effective and why. They need continuing support in development and improvement of data systems. For instance, as strong as Delaware's data system is today, we can benefit from knowledge of cutting edge technology. All states are eager to learn more and to improve the quality of education for our children.

I appreciate the opportunity to address the committee today. Thank you for your leadership. I will be glad to respond to your questions.

Chairman MILLER. Dr. Dougherty?

STATEMENT OF CHRYS DOUGHERTY, PH.D, DIRECTOR OF RESEARCH, NATIONAL CENTER FOR EDUCATIONAL ACCOUNTABILITY

Mr. DOUGHERTY. I would like to thank the first two presenters for making a lot of my points for me.

First, I would agree with Dr. Olson that it is very important to look at growth across the entire achievement spectrum, that it is valuable both for accountability, and it is valuable from the point of view of school improvement.

My organization, the National Center for Educational Accountability, identifies and studies consistently higher-performing schools to see what they do compared to the average performing schools. And looking at student growth is a critical part of this process.

And I would like to thank Dr. Woodruff for emphasizing the importance of longitudinal student data systems at the state level to be able to do these types of models. Our organization has been working very closely as lead partner on the data quality campaign to essentially encourage all states to develop longitudinal student data systems. We have got a packet that should be in your hands that describes a lot of the information about which states have made progress in that area.

Twenty-seven states so far, according to our survey, actually have the critical, three critical data elements in place in order to do, as of next year, a growth model based on longitudinal student data. Now, that doesn't mean that they have every component in place. Dr. Woodruff mentioned assessment system requirements and so forth. But it does mean that from the point of view of building a statewide longitudinal data system they are definitely on track.

And I would like to compliment the Congress for essentially funding longitudinal data grants, which has helped to accelerate this process of states developing longitudinal student data systems. If you had done the same list 3 or 4 years ago, you would have had

fewer than 10 states with the capability longitudinally of doing any kind of growth model. Now it is up to 27. It is very likely it will be over 40 in another 3 years. So that has been very helpful.

I just want to mention that the way growth is handled now as part of AYP and these growth models—and this is reiterating some of the things that have been said—you have got status, which is are enough kids proficient today. You have safe harbor. Are you reducing the percent of kids that are not proficient? And growth is the third.

If kids are way below proficient, are you growing them on a path or a track to proficiency. And Dr. Doran is expert in a lot of the different methods you can use to say what do you mean by on track to proficiency, how do you measure that. It looks like I have a minute left, so I am going to mention a couple of the different ways.

The system Delaware uses is essentially it takes students and it puts them in achievement bands, level one, two, three, four, five. Or California would do far below basic, below basic, basic, proficient, advanced. And basically you monitor the progress of students over the bands. You essentially, as it were, deduct points for kids falling back. You give more points for kids moving forward.

And everybody can understand that. It is very simple. That is called a value table approach. That is one approach.

Another approach is just to draw a trajectory or line between where the kid is now in proficiency. If he is below proficient, it could be a curved line. It could be a straight line. And if you next year are on or above the line, then you are meeting the growth requirement for being on track to proficiency.

And the third approach, which Dr. Doran's organization specializes in, is using statistical models to project or predict whether or not a student will be proficient based on past patterns of students with a certain score in, let's say, 3rd grade and a certain score in 4th grade. What were the odds that that kid would be proficient in 6th grade?

So that uses, again, longitudinal data, which states need to have in order to be able to develop these models and also in order to be able to validate these models to see the extent to which students who were predicted to be proficient actually get there. And that is very critical, the validation part of these growth models.

I finally want to mention that as we move toward putting attention also on kids who are proficient, not only not slipping back below proficiency, but also growing to levels above proficiency, I don't know if the AYP system is the right place to handle that because of the issue of you don't want to offset kids not growing at the bottom end with kids growing at the top end. You don't want to use one to offset the other.

But rather, you want to look at both issues separately and maybe make the growing of the kids at the top end be part of a recognition system. And maybe that is the way to handle it and not through the AYP system.

Thank you very much. I would be happy to answer questions afterwards.

[The statement of Mr. Dougherty follows:]

Prepared Statement of Chrys Dougherty, Ph.D, Director of Research, National Center for Educational Accountability

Mr. Chairman and members of the Committee, I thank you for the opportunity to testify about the use of good educational data, over time, to measure the growth of student achievement. I am Chrys Dougherty, Director of Research at the National Center for Educational Accountability (NCEA), national sponsor of Just for the Kids.

The Center is one of 14 national organizations that are managing partners of the Data Quality Campaign. This campaign is a national, collaborative effort to encourage and support state policymakers to: 1) improve the collection, availability, and use of high-quality education data, and 2) implement state longitudinal data systems to improve student achievement. I will refer in my testimony to the Ten Essential Elements of a statewide longitudinal data system identified by NCEA and the Data Quality Campaign (attached), and to information from NCEA's Survey on State Data Collection which identifies where states are currently in implementing high-quality data systems capable of answering questions critical to improving schools and school systems (a selected list of these questions is also attached).

I have also been privileged to serve on a panel for the U.S. Department of Education's Institute of Education Sciences to review state applications for the state longitudinal data system grant program authorized under title II of the Education Sciences Reform Act of 2002, and currently serve on a panel for the U.S. Department of Education to review state applications to implement growth models for NCLB.

An Overview of Growth Models

"Growth models" can be defined as any analysis or measurement of the progress of individual students over time. The growth models of interest here ask the question: Is the student growing fast enough to be "on track" to reach the desired goal in the desired length of time? For example, is the student progressing well enough to be ready to handle rigorous high school coursework by the time he or she enters high school?

Growth models of this type should be distinguished from conventional "value-added" models, which ask the question, "Is the student growing faster than would be predicted by his or her characteristics?" Typically these characteristics include the student's prior test scores. However, students could be growing faster than predicted for typical students like themselves, and yet not fast enough to reach proficiency in the desired length of time—or ever.

Annual testing in grades 3-8 has been crucial for the development of growth models. These models are based on following students year after year and looking at individual growth every year, rather than waiting several years to find out whether the student has progressed.[1]

Since the desired goal under the No Child Left Behind Act (NCLB) is proficiency, the first question that NCLB growth models address is whether non-proficient students are growing fast enough to reach proficiency in the near future—usually in the next three years.

A second question that NCLB growth models sometimes address is whether already proficient students are growing fast enough to stay proficient.

A third question that these models should address is whether already proficient students are growing to levels higher than proficiency. NCLB as currently written does not encourage states and school districts to address this question.[2] This question is especially important in states where the proficiency standard is below that required to prepare students for college and other postsecondary training for skilled careers.

We would like to encourage school systems to focus on whether students, particularly disadvantaged students, are growing toward readiness for college and skilled careers after high school. Goals and standards that states set for accountability—ones to which sanctions are attached—are likely to be lower than those which school

[1] The ability to look at student growth was a major motivator for the early adoption of grades 3-8 testing in states such as Tennessee, Texas and North Carolina. Annual testing data was critical for Texas's Comparable Improvement growth model, North Carolina's growth model, and Tennessee's value-added model.

[2] The exception to this is NCLB's authorization of funding for Advanced Placement incentive programs.

systems should adopt for purposes of goal-setting, curriculum design, and long-term planning.[3]

Therefore, an incentive for growth to higher levels is probably best accomplished not through the Adequate Yearly Progress (AYP) system, but rather by encouraging the creation of voluntary programs for identifying and publicly recognizing schools that are successful at placing students, particularly disadvantaged students, on a trajectory to these higher standards. Identifying these schools and examining their best practices should be the topic of ongoing research and dissemination.[4]

Data That Is Necessary to Measure Student Academic Growth

The ability to follow individual students over time, as necessary for growth models, requires a longitudinal data system. Specifically, to create growth models, states need at least the following three elements from the list of Ten Essential Elements identified by the Data Quality Campaign (*www.dataqualitycampaign.org*):

- Element One: A statewide student identifier making it possible to follow the same students over time
- Element Three: The ability to link students' test score records over time
- Element Four: Information on untested students and the reasons why they were not tested.

The Status of State Data Systems Capable of Measuring Growth

According to the 2006 NCEA Survey on State Longitudinal Data Systems, 27 states will have the capability of doing a growth model as of the 2007-08 school year, based on their possession of these three elements for at least two years. These states, listed on the Data Quality Campaign's website at *www.dataqualitycampaign.org/survey—results/policy.cfm*, are:

Alaska	Massachusetts	Rhode Island
Colorado	Minnesota	Tennessee
Connecticut	Nebraska	Texas
Delaware	Nevada	Utah
Florida	New Mexico	Vermont
Hawaii	New York	Virginia
Kansas	North Dakota	Washington
Kentucky	Ohio	West Virginia
Louisiana	Pennsylvania	Wisconsin

The Statewide Longitudinal Data System grants have helped many states develop and improve their longitudinal student data systems. These competitive grants from the U.S. Department of Education's Institute of Education Sciences have not only increased the ability of states to do growth models, but also their capacity to provide information to teachers and principals on the academic growth of their students.

Better information is a critical tool for school improvement.

Thank you, Mr. Chairman. I'd be happy to answer any questions you may have.

Essential Elements and Fundamentals of a Longitudinal Data System

While each state's education system is unique, it is clear that there is a set of 10 essential elements that are critical to a longitudinal data system:

1. A unique statewide student identifier that connects student data across key databases across years
2. Student-level enrollment, demographic and program participation information
3. The ability to match individual students' test records from year to year to measure academic growth
4. Information on untested students and the reasons they were not tested
5. A teacher identifier system with the ability to match teachers to students
6. Student-level transcript information, including information on courses completed and grades earned
7. Student-level college readiness test scores
8. Student-level graduation and dropout data
9. The ability to match student records between the P—12 and higher education systems
10. A state data audit system assessing data quality, validity and reliability

---

[3] For a discussion of why accountability standards are often not set high enough to be worthy goals for long-range planning, see "Identifying Appropriate College Readiness Standards for All Students," *www.just4kids.org/en/research—policy/college—career—readiness.*

[4] See *www.just4kids.org* for examples of efforts to identify and recognize higher performing schools and to research and disseminate their practices.

In addition to the 10 essential elements, states need to ensure that they take into account the following fundamental concepts in the construction of their longitudinal systems.

Privacy Protection: One of the critical concepts that should underscore the development of any longitudinal data system is preserving student privacy. An important distinction needs to be made between applying a "unique student identifier" and making "personally identifiable information" available, for example. It is possible to share data that are unique to individual students but that do not allow for the identification of that student. It also is critical to put in place encryption and data security protocols to secure the transmission or transaction of data between and among systems. States should ensure that they bring privacy considerations into the development of each repository and the exploration of each protocol or report.

- Maximizing the Power of Education Data While Ensuring Compliance with Federal Student Privacy Laws: A Guide for Policymakers
- State Longitudinal Data Systems and Student Privacy Protections Under the Family Educational Rights and Privacy Act
- The Family Educational Rights and Privacy Act (FERPA) and State Longitudinal Data Systems
- State Data Systems and Privacy Concerns: Strategies for Balancing Public Interest

Data Architecture: Data architecture defines how data are coded, stored, managed and used. Good data architecture is essential for an effective data system. Many states are in the process of improving their data architecture so that they can clearly communicate with all entities with which they share and from which they receive data. Districts need to know specifically how data elements are defined (e.g., what a "dropout" is), how they should be formatted, and how and when the data should be transferred to the state education agency. Without these standard definitions and dictionaries, state education agencies will have an extremely difficult time making sense of the data received from their districts. With standards in place that are used by everyone, staffing resources and processing or cycle time can be greatly reduced, data can be made available to users when they need them, and reports can be based on clear and common definitions.

Data Warehousing: Many states are in the process of designing and building or upgrading their data warehouses. Policymakers and educators need a data system that not only link student records over time and across databases but also make it easy for users to query those databases and produce standard or customized reports. A data warehouse is, at the least, a repository of data concerning students in the public education system; ideally, it also would include information about educational facilities and curriculum and staff involved in instructional activities, as well as district and school finances. The warehouse should ensure student and teacher confidentiality, allow longitudinal analyses, and include analytical capabilities for its users. Examples of the capabilities that should be available in a data warehouse include, but are not limited to, trend analyses; tracking of students over time and across campuses and/or districts; queries designed and conducted by different users (with different levels of access to detailed data, depending on user classification); and standard summary reports at the campus, district or state level for policymakers and educators. The key to effective data warehousing is the timely and efficient use and reporting of data.

Interoperability: Data interoperability entails the ability of different software systems from different vendors to share information without the need for customized programming or data manipulation by the end user. Interoperability reduces reporting burden, redundancy of data collection, and staff time and resources. It allows for better, faster and clearer reporting of data. It depends on systems having common data standards and definitions. Organizations such as the Schools Interoperability Framework Association work to ensure the creation of platform-independent, vendor neutral open standards that can be used by educators and vendors to design and implement interoperable data systems.

Portability: Data portability is the ability to exchange student transcript information electronically across districts and between P-12 and postsecondary institutions within a state and across states. Portability has at least three advantages: it makes valuable diagnostic information from the academic records of students who move to a new state available to their teachers in a timely manner; it reduces the time and cost of transferring students' high school course transcripts; and it increases the ability of states to distinguish students who transfer to a school in a new state from dropouts. The large interstate movement of students in the wake of Hurricane Katrina made the value of such a system obvious. Data portability is supported by the implementation of interoperable systems, but it requires states that use these systems to have a set of common definitions or protocols.

Professional Development around Data Processes and Use: Building a longitudinal data system requires not only the adoption of key elements outlined in this paper but also the ongoing professional development of the people charged with collecting, storing, analyzing and using the data produced through the new data system. The local school person who inputs course grades needs to understand fully how his/her work fits into the broader data system, the principal needs to understand how data can effect daily school management—both facilities and academic decisions—and policymakers need to understand how their decisions are limited or expanded based on the quality of the data available. For these changes in culture and management to occur, states need to make it a priority to rethink and possibly reorganize how education data is managed throughout the system, increase training and professional development for staff—both managers and users—and assist all employees and stakeholders of the state education system to be active consumers of the longitudinal data system.

Researcher Access: Research using longitudinal student data can be an invaluable guide for improving schools and helping educators learn what works. These data are essential to determining the value-added of schools, programs and specific interventions. States are developing ways to make student-level data available to researchers while protecting the privacy of student records under the Family Education Rights and Privacy Act. Because state education agencies and local school districts usually do not have the resources to conduct this research themselves, providing access to the data to outside researchers with appropriate privacy protections allows critical research to be done at no cost to the state or to school districts.

Policy Implications of State Data Systems in 2006-07

Does your state collect the most relevant data to inform your policy conversations and decisions?

Policymakers and educators need longitudinal data systems capable of providing timely, valid and relevant data. Access to these data gives teachers the information they need to tailor instruction to help each student improve, gives administrators the resources and information to effectively and efficiently manage, and enables policymakers to evaluate which policy initiatives show the best evidence of increasing student achievement.

Does your state have the data to answer these timely questions? Based on responses to the 2006 NCEA survey, only a few states can answer each of these priority questions facing policymakers and educators today.

1. Which schools produce the strongest academic growth for their students? (27 states can answer this question; States must have Elements 1, 3, 4 to answer this question)

Alaska, Colorado, Connecticut, Delaware, Florida, Hawaii, Kansas, Kentucky, Louisiana, Massachusetts, Minnesota, Nebraska, Nevada, New Mexico, New York, North Dakota, Ohio, Pennsylvania, Rhode Island, Tennessee, Texas, Utah, Vermont, Virginia, Washington, West Virginia, Wisconsin

2. What achievement levels in middle school indicate that a student is on track to succeed in rigorous courses in high school? (5 states can answer this question; States must have Elements 1, 3, 6, 7 to answer this question)

Arkansas, Florida, Georgia, Texas, Utah

3. What is each school's graduation rate, according to the 2005 National Governors Association graduation compact? (28 states can answer this question; States must have Elements 1, 2, 8, 10 to answer this question)

Alabama, Alaska, Arizona, Arkansas, Colorado, Connecticut, Delaware, Florida, Iowa, Kansas, Louisiana, Massachusetts, Minnesota, Nevada, New Hampshire, New Mexico, North Dakota, Ohio, Oregon, South Dakota, Texas, Utah, Vermont, Virginia, Washington, West Virginia, Wisconsin, Wyoming

4. What high school performance indicators (e.g., enrollment in rigorous courses or performance on state tests) are the best predictors of students' success in college or the workplace? (4 states can answer this question; States must have Elements 1, 3, 6, 7, 8, 9 to answer this question)

Arkansas, Florida, Georgia, Texas

5. What percentage of high school graduates who go on to college take remedial courses? (14 states can answer this question; States must have Elements 1, 8, 9 to answer this question)

Alabama, Alaska, Arkansas, Florida, Georgia, Hawaii, Louisiana, Massachusetts, North Dakota, Oregon, Texas, Vermont, Washington, Wyoming

6. Which teacher preparation programs produce the graduates whose students have the strongest academic growth? (10 states can answer this question; States must have Elements 1, 3, 4, 5 to answer this question)

Delaware, Florida, Hawaii, Kentucky, Louisiana, New Mexico, Ohio, Tennessee, Utah, West Virginia

State	Policy Questions					
	1	2	3	4	5	6
Alabama	✗	✗	✓	✗	✓	✗
Alaska	✓	✗	✓	✗	✓	✗
Arizona	✗	✗	✓	✗	✗	✗
Arkansas	✗	✓	✓	✓	✓	✗
California	✗	✗	✗	✗	✗	✗
Colorado	✓	✗	✓	✗	✗	✗
Connecticut	✓	✗	✓	✗	✗	✗
Delaware	✓	✗	✓	✗	✗	✓
District of Columbia	✗	✗	✗	✗	✗	✗
Florida	✓	✓	✓	✓	✓	✓
Georgia	✗	✓	✗	✓	✓	✗
Hawaii	✓	✗	✗	✗	✓	✓
Idaho	✗	✗	✗	✗	✗	✗
Illinois	✗	✗	✗	✗	✗	✗
Indiana	✗	✗	✗	✗	✗	✗
Iowa	✗	✗	✓	✗	✗	✗
Kansas	✓	✗	✓	✗	✗	✗
Kentucky	✓	✗	✗	✗	✗	✓
Louisiana	✓	✗	✓	✗	✓	✓
Maine	✗	✗	✗	✗	✗	✗
Maryland	✗	✗	✗	✗	✗	✗
Massachusetts	✓	✗	✓	✗	✓	✗
Michigan	✗	✗	✗	✗	✗	✗
Minnesota	✓	✗	✓	✗	✗	✗
Mississippi	✗	✗	✗	✗	✗	✗
Missouri	✗	✗	✗	✗	✗	✗
Montana	✗	✗	✗	✗	✗	✗
Nebraska	✓	✗	✗	✗	✗	✗
Nevada	✓	✗	✓	✗	✗	✗
New Hampshire	✗	✗	✓	✗	✗	✗
New Jersey	✗	✗	✗	✗	✗	✗
New Mexico	✓	✗	✓	✗	✗	✓
New York	✓	✗	✗	✗	✗	✗
North Carolina	✗	✗	✗	✗	✗	✗
North Dakota	✓	✗	✓	✗	✓	✗
Ohio	✓	✗	✓	✗	✗	✓
Oklahoma	✗	✗	✗	✗	✗	✗
Oregon	✗	✗	✓	✗	✓	✗
Pennsylvania	✓	✗	✗	✗	✗	✗
Rhode Island	✓	✗	✗	✗	✗	✗
South Carolina	✗	✗	✗	✗	✗	✗
South Dakota	✗	✗	✓	✗	✗	✗
Tennessee	✓	✗	✗	✗	✗	✓
Texas	✓	✓	✓	✓	✓	✗
Utah	✓	✓	✓	✗	✗	✓
Vermont	✓	✗	✓	✗	✓	✗
Virginia	✓	✗	✓	✗	✗	✗
Washington	✓	✗	✓	✗	✓	✗
West Virginia	✓	✗	✓	✗	✗	✓
Wisconsin	✓	✗	✓	✗	✗	✗
Wyoming	✗	✗	✓	✗	✓	✗

Chairman MILLER. Mr. McWalters?

STATEMENT OF PETER MCWALTERS, COMMISSIONER OF ELEMENTARY AND SECONDARY EDUCATION, STATE OF RHODE ISLAND

Mr. MCWALTERS. Chairman Miller, Ranking Member McKeon, thank you for this opportunity. My name is Peter McWalters. I am the commissioner from Rhode Island. I have been there for 15 years. And before that, I was a superintendent of schools in Rochester. I am clearly an urban educator.

I am pleased to be able to talk to you today as you consider reauthorizing No Child Left Behind. I was the president of CCSSO in 2000, 2001 when we authorized this. And I supported it then. I support it now.

It represents the very best form of federal intent. It essentially is the Civil Rights bill. It is part of a children's bill of rights. And it pushed states to focus on success for every student.

The emphasis of standards and assessments and accountability on public information was needed then as it is now. And it has been beneficial for the nation. Now 5 years down the road, I think we can see some areas in which the law could and should be modified to help achieve the goals we all share.

As CCSSO has said in its recent recommendations regarding No Child Left Behind reauthorization, we are in a new stage of standards-based reform. Many of the basic foundational pieces are in place. The question now is, how do we build on the use of these foundations to improve student achievement and close the gaps?

I would submit to you that this will require innovation, change beyond currently understood, capacity building and retooling of systems, and quite honestly, judgments that are based on leadership, content, capacity, and the context of districts and schools. We need a federal law that values these things.

As you prepare to reauthorize No Child Left Behind, I ask you to consider three issues: how states determine whether schools have met their targets, how we publicly identify schools that have missed their targets, and how states can best deliver assistance and implement consequences to help districts as well as schools meet these goals.

As you know, schools may be identified for improvement if they miss any single one multiple target established in the law. And these targets are almost exclusively based in the tests that states administer at seven grade levels.

We are not afraid to use student performance as the ultimate measure of school improvement. Our testing system in Rhode Island developed with the support of the federal funds is a tri-state partnership under which Rhode Island, New Hampshire, and Vermont established a common set of grade-level expectations and standards and developed an assessment system lined up with those standards. This partnership, known as the New England Common Assessment Program, is exactly the type of initiative that the federal government should continue to support.

In addition to our state assessment system, we believe in Rhode Island in a number of means by which we can and do measure school performance. We administer attitudinal surveys to students,

parents, teachers, and administrators. We visit schools on a structured visitation. We publish the results of both of those—they are all online.

Parents get access to all this information. We measure school climate as in safety. We measure student connectedness. Does anybody here know me? Is anybody listening to me?

We measure instructional leadership. We measure instructional practice, teacher competencies, as in, do they even know what the standards are. We track all of this stuff as well as parent involvement. We conduct peer review visits at every school.

Every school is required by law to write an annual school improvement plan and submit district plans to us. And if you are in intervention, we get to not only review them, but we approve them. We have a very aggressive statute of progressive support and intervention.

Test results should be the initial measure of districts and schools. But the law should allow states to employ indicators in addition to student performance to determine whether schools and districts are making adequate yearly progress.

These indicators should include measures of capacity such as school climate, teacher expectations, leadership, instructional leadership, teacher development, program implementation fidelity, and parent engagement. These indicators should be supplemental to assessment results, but they should be allowed to be part of an overall determination of school as well district progress.

As you know, NCLB is quite prescriptive in regards to identifying schools and districts that have missed annual targets. Under the terms of the law, all schools that missed even one target are placed in the same status: identified for improvement. This label tells us only that the school has not met the target. It does not tell us why.

I have seen the school fail in 1 year go from high-performing to insufficient progress because it missed a single target. And we find this hard to explain in terms of the public policy or what Valerie called face validity.

I believe that the law should establish a graduated system of classifications for schools and districts that have identified for improvement. The identification of schools and districts should include information as to how many targets were missed as well as over how many years. The identification of schools and districts should also indicate the capacity of the school or district to meet these targets as determined by indicators other than test results.

Finally, I ask you to consider how states develop support systems and intervention strategies for schools and districts that have been identified for improvement. We don't need an intervention system that is based on a score card. We need a system that will give us multiple ways to measure all the components of the viability of a school in a district and to offer scaffolded responses based on the needs of schools and districts.

The system as it stands is not designed to give schools a blueprint for success. It is a retributive system. We will not shrink from our responsibility of raising achievement and closing the gap. But we need the law to value our experience and leverage the expertise

and give us more options over schools that are identified for improvement.

Not all schools admit their targets are in the same place. Some may be truly dysfunctional institutions in need of a great deal of help, even restructuring. Others may be on task and the path toward success. How do states know if this is the case? Only through multiple measures.

Indicators of measures of leadership, instructional leadership capacity, school climate, community involvement, and program integrity. Only through this can we determine the course, the appropriate course of action to take.

Now that we are 5 years into the implementation of this law, it is obvious that many schools that have missed their annual targets are doing all they can within failing systems. That is, school improvement is often a matter of district capacity. In these instances, the intervention at school level will do nothing to solve the underlying systemic problem.

When a state intervenes in a school that has missed targets, the state must have on-hand the complete picture of the school and district capacity. The law should not prescribe our responses. It should give us the authority to use our best professional judgment to build school improvements.

The Rhode Island approach has been entered into district negotiated agreements that we write, negotiate, and finally approve on a program, budget, and personnel basis. That is pretty powerful. This is part of our process of progressive support and intervention.

We are ready to do the work. To do that, we need an NCLB that is more than just a score card based on student performance and a list of mandated responses. We need indicators to measure all components of the health and capacity of the system.

Chairman MILLER. Mr. McWalters, I am going to ask you to wrap up.

Mr. MCWALTERS. Very good.

The last piece that I would say is when you passed this authorization, there was a sense of impatience on your part, which was well-deserved at that time. I think 5 years in the credibility of individual states' capacity is now known and can be reviewed in a peer review system.

[The statement of Mr. McWalters follows:]

Prepared Statement of Peter McWalters, Commissioner of Elementary and Secondary Education, State of Rhode Island

Chairman Miller, Ranking Member McKeon, and members of the Committee, thank you for the opportunity to testify today on improving the ways we measure student progress. My name is Peter McWalters, and I am the Commissioner of Elementary and Secondary Education in the State of Rhode Island, where I have served for 15 years. I am also a past-president of the Council of Chief State School Officers and a former Superintendent of Schools in an urban district, Rochester, New York.

I am pleased to be able to talk with you today as you consider reauthorization of the No Child Left Behind Act. I supported the law in its passage. It represents the best form of federal intent and has pushed the states to focus on success for every student. The emphasis on standards and assessments and on public information was needed at the time, and it has been beneficial to the nation. But now, five years down the road, I think we can see some areas in which the law could and should be modified to help us achieve the goals that we all share.

As CCSSO has said in its recent recommendations regarding NCLB reauthorization, we are in a new stage of standards-based reform. Many of the basic foundations are in place. The question now is: How do we build on and use these foundations to improve student achievement and close achievement gaps? I would submit to you that this will require innovation, capacity, and judgments that are based on district capacity to respond to specific conditions that have led to low student achievement. We need a federal law that values those things.

As you prepare to reauthorize NCLB, I ask you reconsider three issues:

- how states determine whether schools have met their targets,
- how we publicly identify schools that have missed their targets, and
- how states can best deliver assistance and implement consequences to help schools meet their goals.

As you know, schools may be identified for improvement if they miss any single one of the multiple targets established in the law. And these targets are almost exclusively based on the tests that states administer at seven grade levels.

We are not afraid to use student performance as the ultimate measure of school improvement. Our testing system in Rhode Island, developed with the support of federal funds, is a tristate partnership, under which Rhode Island, New Hampshire, and Vermont established in common a set of grade-level standards and expectations and developed an assessment system lined up with those standards. This partnership, known as the New England Common Assessment Program, is exactly the type of initiative that the Federal government should continue to support.

In addition to our state assessment system, we have in Rhode Island a number of means by which we can—and do—measure school performance. We administer an annual survey to all students, teachers, and parents, and from the results of this SALT Survey we tabulate "Learning Support Indicators" that measure school climate, instructional practices, and parental involvement. We conduct peer-review visits at every school in the state every five years. Each school is required by law to write an annual School Improvement Plan, and each district writes an annual District Strategic Plan, and these plans are at the center of our work with all schools and districts.

Test results should be the initial measure of the school. But the law should allow states to employ indicators in addition to student performance to determine whether schools and districts are making Adequate Yearly Progress. These indicators could include measures of capacity such as evaluations of school climate, instructional practices, instructional leadership, teacher development, program implementation, and parental engagement. These indicators should be supplementary to assessment results, but they should be allowed as part of the overall determination of school and district progress.

As you know, the NCLB is quite prescriptive in regard to identifying schools and districts that have missed annual targets. Under the terms of the law, all schools that miss even one target are placed in the same status: Identified for Improvement. This label tells us only that the school has failed; it does not tell us why. I have seen a school fall in one year from high performing to insufficient progress because it missed a single target, and we find this hard to explain to the school and to the public at large.

I believe that the law should establish a graduated system of classifications for schools and districts that have been identified for improvement. The identification of schools and districts should include information as to how many targets were missed as well as for how many years. The identification of schools and districts should also indicate the capacity of the school or district to meet all targets, as determined by indicators other than test results.

Finally, I ask you to reconsider how states develop support systems and intervention strategies for schools and districts that have been identified for improvement. We don't need an intervention system that is based on a scorecard. We need a system that will give us multiple ways to measure all components of the health and the capacity of schools and districts and to offer scaffolded responses based on the needs of the school or district. The system as it stands is not designed to give schools a blueprint for success. It is a retributive system.

We will not shirk our responsibility for raising achievement and closing the achievement gap. But we need the law to value our experience and expertise and give us more options once schools are identified for improvement. Not all schools that miss their targets are in the same condition. Some may be truly dysfunctional institutions in need of a great deal of help—even restructuring. Others may be on task and on the path toward success. How do states know if this is the case? Only through multiple measures—indicators to measure leadership, instructional capacity, school climate, community involvement—can we determine what course to take to help schools meet their goals.

Now that we are five years into implementation of the law, it is obvious that many schools that have missed their annual targets are doing all that they can do within a failing system. That is, school improvement is often a matter of district capacity. In these cases, state intervention at the school level will do nothing to solve the underlying systemic problems.

When a state intervenes in a school that has missed targets, the state must have on hand a complete picture of the school and district capacities. The law should not prescribe our responses. It should give us the authority to use our best professional judgment to build school improvement. The Rhode Island approach has been to enter into District Negotiated Agreements on program, budget, and personnel with those districts that have missed their annual targets. This is part of our process of Progressive Support & Intervention, which is based on multiple indicators that present information for broader and deeper than assessment results.

We are ready to do the work. To do that, we need from NCLB more than just a scorecard based on student performance and a list of mandated responses. We need indicators to measure all components of the health and capacity of the system. We need intervention strategies that help us build the capacity in each identified school and district. And we need the freedom and capacity to do our work, while always keeping the goals clear and the actions and outcomes transparent so as to improve the public-education system.

I ask, therefore, that you consider revising the prescribed sequence of mandated responses to Title I schools that have been identified for improvement so that states can develop graduated support and intervention strategies that best meet the needs of each identified school.

I have asked you today for a good deal of accountability at the state level, for I believe that the states have the ability to take on this challenge. When Congress passed and the President authorized the NCLB, there was a general sense of impatience with progress that the states had made. The law is therefore both comprehensive and prescriptive in regard to state responsibilities. The states have taken on these responsibilities in a serious and committed manner, and I therefore believe we are ready to move to a new level of shared understanding. States should be able to submit their annual compliance plans, which the Education Department would verify and accept after good-faith peer review

The CCSSO recommendations for NCLB reauthorization include several items that support the points I have brought to you today, including calling on Congress to allow states to include additional relevant data in making judgments about school progress, allowing states to differentiate consequences for schools that have missed their annual targets, investing more in state capacity to assist and intervene in districts and schools that have missed their targets, and creating a new process for innovative models and a greatly revised system of peer review that would allow states to continuously innovate in accountability and other areas—with proper guarantees for results.

Thank you for your attention and leadership on these important issues. I have with me several supportive documents regarding the accountability system in Rhode Island that I would like to present to you for your records, and I look forward to any questions you may have.

Chairman MILLER. Thank you.

Dr. Doran?

STATEMENT OF HAROLD C. DORAN, SENIOR RESEARCH SCIENTIST, AMERICAN INSTITUTES FOR RESEARCH

Mr. DORAN. Thank you. Chairman Miller, Ranking Member McKeon, and honorable members of the committee, thank you for this opportunity to share my thoughts on ways to improve the No Child Left Behind Act.

My name is Harold Doran, and I am a senior research scientist at the American Institutes for Research in Washington, D.C. In this role, I help states and districts across the country develop their testing and accountability systems. I am also a former classroom teacher and elementary school principal in Tucson, Arizona.

The question I have been asked to respond to today is whether the AYP provisions would benefit from having additional ways to

evaluate schools, what some refer to as multiple measures, and whether these measures can be joined to form a compensatory accountability system. The term "compensatory" denotes that not meeting AYP under one measure could be compensated for using a secondary measure. I believe the provisions could be strengthened if multiple measures were added.

In my discussion today, I would like to explain this position and suggest specific measures that I believe would strengthen the legislation. I emphatically support the use of multiple measures, as do most educational experts. However, there are multiple views on what set of measures to be included in accountability systems. Even more challenging is how these measures can be combined in forming a compensatory accountability system.

To reduce ambiguity, I would offer the following definition of multiple measures for today's conversation: an accountability system that includes multiple measures uses test scores from more than a single test, achievement indicators collected by other means, or various statistical methods for evaluating the data. By this definition, NCLB already uses multiple measures.

But the law does not permit for one to compensate for another measure. I believe the integrity of the law would be enhanced if it were modified to accommodate the following: multiple measures; and allow states to use those measures to create rigorous compensatory systems.

First, any consideration of new measures, however, must first be met with a discussion of criteria to avoid watering down any of the current systems. One, including new indicators should result only in added rigor to core content areas. Two, incorporating multiple measures should not result in systems that are too complex so that they are difficult to implement or confusing to parents and educators.

I have four specific recommendations. Two of these recommendations would add measures that could serve in a compensatory role. One recommendation adds to AYP. And the last is a recommendation to ensure system integrity.

NCLB currently monitors the proficiency rates of high school students in language arts, reading and math. When students do not reach levels of proficiency on the statewide regular tests, their only option is to retake the same test a year later.

However, an alternative that could be used is to provide students with an opportunity to enroll in targeted coursework that targets their specific area of need and allow for them to pass an end-of-course examine that allows for them to demonstrate mastery of the content.

For instance, a student may not reach proficiency on the statewide test because it were known that he struggled with concepts in geometry. Subsequently, the student could enroll in a geometry course, demonstrate proficiency via a new state-developed end-of-course exam that is equally as rigorous as the statewide NCLB test.

Learning is fundamentally about change. However, the methods by which AYP are currently calculated do not follow this logic and, in many ways, are actually biased. The current reality is that the

mathematical model used to measure proficiency rates must be improved.

For example, a school with many students scoring in the highest performance category can have a drop in students' academic performance that still remains above proficiency and still be classified as a school making AYP. In contrast, a school with many students beginning well below proficiency and learning at remarkable rates, is likely not to be recognized as a high-performing school.

It is my recommendation that AYP calculations include results obtained from growth models as another method for evaluating schools. NCLB currently requires students to participate in science assessments beginning in 2008. However, the results of those assessments will not be included currently in AYP calculations. It is my recommendation that they should be. It is also possible to develop end-of-course exams in science, as previously suggested.

Last, I would like to offer a suggestion on the use of NAEP. It cannot be used to measure AYP, but it can be used to inform how state performance standards are set and partly used to determine overall system integrity. I would like to recommend that this committee support a research agenda that would investigate and report how best to establish links between NAEP and the various state assessment programs.

In many respects, the variability in standards and difficulty of the assessment programs across states is important and reflects idiosyncrasies in the educational programs. On the other hand, this variability presents a significant challenge, given that we live in a highly mobile society.

It is my view that reauthorized versions of NCLB should establish national policy using NAEP to illustrate the comparability of proficiency levels across the country. This information would be extremely valuable as states build or refine their standards and assessment programs. It will also provide policymakers with a window to assess system integrity.

Thank you for your time. I hope these suggestions are helpful. And I am grateful to answer any questions that you may have.

[The statement of Mr. Doran follows:]

Prepared Statement of Harold C. Doran, Senior Research Scientist, American Institutes for Research

Chairman Miller, ranking member McKeon, and honorable members of the committee, thank you for this opportunity to share my thoughts on ways to improve the No Child Left Behind Act. My name is Harold Doran, and I am a senior research scientist at the American Institutes for Research (AIR) in Washington, DC. In this role, I help states and districts across the country develop their testing and accountability systems.

The question I have been asked to respond to is whether the adequate yearly progress (AYP) provisions in NCLB would benefit from having additional ways to evaluate schools, what some refer to as multiple measures, and whether these measures can be joined to form a compensatory accountability system. The term compensatory denotes that not meeting AYP under one measure could be compensated for using a secondary measure.

I believe the AYP provisions could be strengthened if multiple measures were added. In my discussion today, I would like to explain this position and suggest specific measures that I believe would strengthen the legislation.

Why Multiple Measures?

I emphatically support the use of multiple measures, as do most educational experts. However, there are multiple views on what set of measures to include in ac-

countability systems. Even more challenging is how these measures can be combined in forming a compensatory accountability design. To reduce ambiguity, I would offer the following definition of multiple measures for today's conversation: An accountability system that includes multiple measures uses test scores from more than a single test, achievement indicators collected by other means, or various statistical methods for evaluating the data.

By this definition, NCLB already relies on multiple measures. But the law does not permit one measure to compensate for another measure. I believe the integrity and strength of the law would be enhanced if it were modified to accommodate the following:

1. Permit for multiple measures; and
2. Allow states to use those measures to create rigorous compensatory systems.

Any consideration of new measures, however, must first be met with a discussion of criteria to avoid watering down our current systems:

1. Increased Rigor. Including new indicators should result only in added rigor to core content areas.
2. Simplicity and Transparency. Incorporating multiple measures should not result in complex systems that are difficult to implement or that are confusing to parents and educators. The elegance of simplicity, combined with a focus on rigor, will guard against over-engineering accountability designs.

Specific Recommendations for Multiple Measures

I have four specific recommendations. Two of these recommendations would add measures that could serve in a compensatory role, one recommendation adds to AYP, and the last is a recommendation to ensure system integrity.

End-of-Course Exams

NCLB currently monitors the proficiency rates of high-school students in language arts/reading and math. When students do not reach levels of proficiency on the statewide regular tests, their only option in many cases is to retake the same test. However, an alternative that could be used is to provide students with an opportunity to enroll in coursework that targets their specific areas of need and allow for them to pass an end-of-course test that demonstrates mastery of the content.

For instance, a student may not reach proficiency on the statewide NCLB test only because he struggles with concepts in geometry. Subsequently, the student could enroll in a geometry course and, at the end of this course, demonstrate proficiency via a state-developed end-of-course exam in geometry that is equally as rigorous as the statewide NCLB test.

Growth Models

Learning is fundamentally about change. However, the methods by which AYP are currently calculated do not follow this logic and are, in many ways, biased.

The current reality is that the mathematical model used to measure proficiency rates must be improved. For example, a school with many students scoring in the highest performance category can have a drop in students' academic performance that still remains above the proficiency bar and still be classified as making AYP. In contrast, a school with many students beginning well below proficiency, but learning at remarkable rates, is likely not to be recognized as a high-performing school.

It is my recommendation that AYP calculations include results obtained from growth models as another method for evaluating schools. The results from these models can be used in a manner similar to the safe-harbor provisions as another way to make AYP. If permitted, the models must conform to the same high expectations for proficiency as currently required and not simply reward growth.

Incorporate Science Results into AYP

The 2001 NCLB requires students to participate in science assessments beginning in 2008. However, the results of those science assessments are not included in the current AYP calculations.

Including science in AYP calculations will encourage schools to emphasize science as a component of their core curricula. It will also be possible to develop end-of-course exams in science as previously suggested.

National Assessment of Education Progress (NAEP) Research for Comparability

Last, I would like to offer a suggestion on the use of NAEP—it cannot be used to measure AYP, but it can be used to inform how state performance standards are set and partly used to determine overall system integrity. I would like to recommend that this committee support a research agenda that would investigate and

report how best to establish links between NAEP and the various state assessment programs across the country.

In many respects, the variability in content standards and difficulty of the assessments across states is important and reflects critical idiosyncrasies in the educational programs. On the other hand, this variability presents a significant challenge given that we live in a highly mobile society. For example, a student attaining mathematical proficiency in Arizona may attend college and/or obtain professional work outside of that state.

Hence, my view is that reauthorized versions of NCLB should establish national policy using NAEP to illustrate the comparability of proficiency levels across the country. This information would be extremely valuable as states build and/or refine their standards and assessment programs. It will also provide policymakers with a window to assess system integrity.

Should the committee accept the notion that additional indicators are necessary to establish more robust systems, I would then encourage the committee to further consider how these multiple indicators can be combined to form a judgment about school quality that still aligns with the basic tenets of proficiency set forth in the legislation.

I hope these suggestions are helpful as this committee moves forward with deliberations related to NCLB improvements. I am grateful for the opportunity to testify today and am happy to answer any questions you may have.

Chairman MILLER. Thank you all very much for your time.

Dr. Doran, you say on the bottom of the first page of your statement that by definition NCLB already relies on multiple measures, but the law does not permit one measure to compensate for another measure.

And, Commissioner McWalters, you said in your statement that these indicators should be supplementary to assessment results, but they should be allowed to be as part of an overall determination of school and district progress. Are those two things consistent?

Mr. DORAN. Maybe I can clarify exactly what I mean and explore for just a moment. Currently there is a limited set of multiple measures that are permitted. In reality, a student has a single opportunity to demonstrate proficiency on the test. We know that tests when designed well can provide very useful and good information. But the reality is some kids have mastered the content, but for one reason or another, didn't have an opportunity to demonstrate their proficiency on the test when it was given on that day.

And I think what I am saying is this. You talk to practitioners. You talk to statisticians. You talk to testing professionals. And we might say that if I had a different day or a different way for a student to demonstrate their proficiency on this content, he would have. I know the student has mastered the material. But it just didn't work today. So I need a different day to test, or I need a different way. The goal is still the same: evaluate whether the student has mastered the concept. Just provide multiple, parallel tracks to identify whether the student has done so.

Chairman MILLER. Commissioner McWalters?

Mr. MCWALTERS. I would concur with that. There are two different multiple measures we are talking about. One is actually about student performance. And the other is capacity issues. I meant don't let the capacity issue measure. The indicators of whether they are on task should not somehow compensate for student performance. But I would concur.

We are a state that very much is trying now to come up with embedded assessments that can be audited for reliability and use to drive practice. Those kind of measures when done right ought to be compensatory as in added to and part of an explanation.

Chairman MILLER. Let me follow up on what you just mentioned. Because in your testimony, you also stated—and this is what concerns me—"Now that we are 5 years in the implementation of the law, it is obvious that many schools that have missed their annual targets are doing all that they can do within a failing system."

Mr. MCWALTERS. Right.

Chairman MILLER. And I think all of us here as we have visited schools and schools that haven't made AYP and they show you what changes they are making, you leave that school and get in your car and drive away thinking they don't have a chance.

Mr. MCWALTERS. Right, right. That is right.

Chairman MILLER. Because you just don't see any change in the capacity to do what is necessary. They have moved everybody around. They have given people titles.

Mr. MCWALTERS. Right.

Chairman MILLER. But it is just not going to happen. And it hasn't happened for the last 20 years in the same schools.

Mr. MCWALTERS. Right.

Chairman MILLER. So you start to think, you know—and so, I am intrigued with the idea of multiple indicators as also being able to give you a handle on what is going on in that school or even within that district.

Mr. MCWALTERS. Yes, right.

Chairman MILLER. But certainly, within that school of whether it is time for professional development or teachers to work together or to review one another's activity, all of these things that we think measure a learning environment. But again, you are really talking about two separate purposes.

Mr. MCWALTERS. That is right, sir.

Chairman MILLER. Is that correct?

Mr. MCWALTERS. Absolutely. I think most of us—again, when I said I was an urban educator, I was in a state that didn't have an urban capacity. The state could not intervene at the district level because some of our urbans they are bigger institutions almost than the State Department.

So when this law started—and I think it started in the right place—all commissioners were into school improvements where you can go in and you can possibly restructure a school to work for a while. But if you step back and that is in a system that is dysfunctional, then that system will eventually come back to neutral, if you will.

So this issue of what other measures of school health, district health aligned from state health to school room is very complicated business. And it is only actually with the emerging information systems that you can start tracking expenditures, time on task, teacher development.

And the most impenetrable ones so far is when you have bad teacher practice with kids who have never been given a fair chance. And then you find teachers who actually begin to understand their own limitations when they see good standards and good feedback.

And then you start realizing teacher retooling is part of an enormous investment strategy.

I think my only point is if you don't know all of that and you just keep using one indicator, I just don't see the viability of that changing the improvement structure.

Chairman MILLER. Dr. Dougherty, do you want to comment?

Mr. DOUGHERTY. I would like to add that I am hearing two issues here. One is a more nuanced way of determining whether the kid is okay. Is the kid on a trajectory to being proficient. And the second is a more nuanced approach to whether the school is okay and the school is on a trajectory.

Chairman MILLER. And a district.

Mr. DOUGHERTY. And a district. And that is a very important point because school systems—schools exist within systems. And a lot of times the problem is the system is dysfunctional.

Chairman MILLER. Okay.

Secretary Woodruff, let me ask you this. This 2007 school year you are going to be using an approved growth model. Is that correct?

Ms. WOODRUFF. Yes, sir.

Chairman MILLER. What is the biggest change that you think you are going to notice?

Ms. WOODRUFF. We don't know. We will be calculating the school's rating based on the traditional model. We will also be calculating using the growth model. And then we will be able to see whether or not there is any difference in the school rating between the two. So until we actually implement and evaluate that implementation, I really can't give you a clear answer.

Chairman MILLER. Thank you. My time is up.

Mr. Olson, I am going to have to get you on a second round here. But I am quite intrigued with your track record in terms of administering these adaptive tests. And I would like to come back to that.

But I would like now to recognize Congressman McKeon.

Mr. MCKEON. Thank you, Mr. Chairman.

And just following up a little bit on the line of questioning that you were doing with Commissioner McWalters and Dr. Doran, you are talking about having a dual type different modes of testing because it would do a better job.

One of the things that I have found in talking to people is their complaint already of having too many tests. Would this be another layer on top of that that they would have to deal with?

Mr. DORAN. A couple of issues.

One, in thinking about this a bit, I think it is clear that we are talking about two buckets, two kinds of things that we want to collect indicators on and about schools. School process indicators, things that are illustrative of how healthy the school is in its instructional leadership and how well students are spending time on task and so forth. And those don't necessarily become quantifiable in the sense of whether students have mastered the core content or not. So those are the school process variables bucket. And those are extremely important.

Then there is the other bucket, which are those measures that are designed to specifically measure whether students have met the outcomes that are expected of them or not. Now, with respect

to student outcomes, if we had multiple measures—that is, other ways that we could evaluate whether the students have mastered the contents or not—I wouldn't necessarily suggest that students would be tested multiple times, per say.

I think that students should be given multiple opportunities to demonstrate the mastery of the concept. So for example, if a student did fine and demonstrated their mastery of the concept on the regular statewide assessment, that is fine. That is the only assessment maybe that student would need to participate in unless the school or the classroom teacher for other reasons wanted the student to participate in something else.

However, if the student didn't demonstrate mastery of the concepts on that particular test, I think there should be multiple avenues from which the school—the state has designed a system such that the school can choose an alternative path. Now, I don't think that, based on my conversations with professionals and state departments of education, my experiences as a practitioner, which was 10 years, that people would push back and reject an opportunity to allow students to have multiple opportunities to demonstrate their mastery of the concept.

I think where people would push back is if students were required to participate in repetitive tests that didn't give them useful information upon which they could make instructional diagnoses from there.

Mr. MCWALTERS. I would completely concur with that. When I talk about multiple assessments, I think one of the things that we are still missing is that the test is perceived as a state test. And thank God, now I think most of us have at least got standards in systems where they are aligned. But teachers don't own them yet.

And until we have worked at the level of teachers developing assessments just like the state tests or versions of it that I would call embedded, much more performance-based, much more on demand and that the state's obligation is to have a system that is auditing that so it is either got quality and it is reliable.

But any of you that know anything about the writing process strategy statewide or nationally know that it is hard work and it is probably extensive to get it embedded. But until teachers begin to own the assessment decisions that would add up to improving the state test, then you are still doing a dip stick strategy and you are not going to change practice substantively.

So the teachers I talk to don't think of my instrumentation as additional testing. They think of it as instructional assessment.

Mr. MCKEON. But it still takes more time away from classroom instruction because they have to do another——

Mr. MCWALTERS. The ones that I am talking about would be done right in an instructional program. It would be part of the instructional practice just like a quiz is today.

Mr. MCKEON. Okay.

Mr. MCWALTERS. If you know what I mean.

Mr. MCKEON. Okay. But a quiz also takes time away from instruction. I mean, at some point whenever you are evaluating, you are taking time away from instruction. I am just saying that was one of the complaints we have is we already have all these tests.

And I am not saying anything about the validity of it, the importance of it.

Just, I think, when you say we get push back on some things, you get push back on just about everything.

Secretary, one of the questions I had is we both come from the largest state. You are one of the smaller states. Do you think what you are doing could be replicated with the number of districts we have, the number of schools we have within our state and then the same thing across the country?

Ms. WOODRUFF. Well, actually, the growth model that we have in place absolutely could be used in small states, large states. It really doesn't matter. We are using the value table. It is very simple to understand, as Dr. Dougherty mentioned. Students are given points for different progressions toward proficiency. If the student slips, then the school gets fewer points. So the system itself is one that really can be used in a small system or a large one. That is not an issue at all.

Could I comment on the previous conversation for a moment?

Mr. MCKEON. Go ahead.

Ms. WOODRUFF. One of the things that I think that we have gotten away from is helping teachers and others understand that assessment has been and always will be a part of instruction and that those quizzes and end of course assessments and so forth are important. In Delaware, we have a student accountability system. And at certain levels students who are well below our standard must attend summer school.

We have developed a system by which school districts can bring to us what we call other indicators of performance. And if students can show proficiency according to those other indicators, then they do not have to go to summer school and face other consequences like not going to the next grade and so forth.

So I think that what both Dr. Doran and Commissioner McWalters are talking about in terms of other kinds of assessments really can be done. The system we have now probably isn't as sophisticated as it ought to be. But something like that makes sense to families and makes sense to students as they get older certainly and certainly to teachers.

Mr. MCKEON. Thank you.

Chairman MILLER. Mr. Kildee?

Mr. KILDEE. Thank you, Mr. Chairman.

Commissioner McWalters, you testified in support of a differentiated interventions for schools that do not meet AYP, depending on how close they are. Can you describe how you might differentiate the consequences for schools that fall short a little, fall short a lot?

Mr. MCWALTERS. Well, right now our practice—we are actually in this practice. We have gone out—we have systems, and they tend to be embedded in big urban systems. I am going to be dramatic.

You have flat line indicators. I mean, the first indication is teach, for God's sake. And that is a pretty heavy assessment. But when you go in, usually when you find there is a pretty complicated set of dysfunctions from leadership to school culture, attitudes. I mean,

you just want to shut the place down, which that is the one dramatic thing we can do.

But the truth is you go from there to places that have reasonably good cultures, but they just internalize low expectations. They love the kids, but they are not working with them. So you need to know that when you are going in there. And you need to know whether or not it is about alignment, time on task, command, control.

You need to settle either those initiatives at the state and district level. And once you have them in your tool kit, you need to know whether the district is part of that problem. Is it the districts that have the systems of dysfunction? And if it does, that changes the trajectory of change.

When I talk about AYP, I have two images. One is a realistic one for a school and a realistic one for a big system with a series of alignments that all have to be dealt with. So I think my point is I am in a little state that has enough information systems on health, time, expense, personnel that that is the level of intervention that we are now dealing with.

And I just see differentiated treatments for different schools. There is a phrase in my state now, "Great schools look awfully similar. Terrible ones can look awfully different."

Mr. KILDEE. Well, let us take this. You have a school A and school B. One just barely missed AYP. And the other one just was way, way down the scale.

Mr. MCWALTERS. Right. That is right.

Mr. KILDEE. Can't we have effects, penalties, consequences, whatever you want to call them?

Mr. MCWALTERS. That is right.

Mr. KILDEE. Do you apply those effects, consequences, penalties differently in those instances?

Mr. MCWALTERS. Well, I would like to be able to—yes, my answer is I think we have to have better degrees of judgments made about what the intervention and the penalties are. And I think those should be in a proposal that is kind of a change theory or status that is reviewed by a peer review structure so that it is not hidden, it is not made up on the spot. It is a whole program of that is one reviewed.

Because one of the other issues I think we have to admit is we are at a scale of intervention that is still an experiment in 50 states. None of us have an answer here. I need both assurance and cover that in good faith I am doing public policy work that can be tracked over time for its effectiveness. And I think that is what the peer review system ought to kind of review and sanction.

Mr. KILDEE. Well, for example, at one point you might require tutoring for students.

Mr. MCWALTERS. Right.

Mr. KILDEE. Because perhaps there was a great differential between where they should be. One just barely fell short. Is there something short of tutoring one could do in that school that would help raise that?

Mr. MCWALTERS. Well, I will give the example of a—in the first round, I think the drama was needed because it uncovered those places we are hiding behind averages. But once you got at that, many of the places actually got on task, identified through

disaggregation what they had to do, and they went about the business of doing it.

But now that we are into this over time, you have schools that kind of drop in and drop out. And to go in there effectively, sometimes they see it coming. Sometimes it is as simple as a cohort question. You want to be able to go in with an instrumentation.

Sometimes it is instructional practice. Sometimes there was a change in leadership. And sometimes it is more time on task like tutoring. I am suggesting that all of those are decisions that need to be made in the context of a really comprehensive assessment of where the school or district is.

Mr. KILDEE. Thank you. Just one more question. Suppose one of these groups whom we disaggregate the data for falls short and that could bring the whole school out of compliance with AYP.

Mr. MCWALTERS. Right.

Mr. KILDEE. Is there something we can do rather than say that school is out of AYP and therefore must suffer the consequences, the effect, whatever you want to call it, that we do something for that one group to help raise them up? Or do we just declare the whole school not achieving AYP?

Mr. MCWALTERS. I would say you just ask me. That is the biggest question in my state now on the periphery. When you have a system that is perceived to be a pretty good system, good system, good school, in one indicator, usually second language or minority or poor kids in a system that they are a tiny percentage, in those early days, that was exactly what I needed because you could go after people that never talked about it.

But now that everybody knows that is the indicator, once you have that, this issue of saying the school is now not in AYP and is in need of improvement it is—I don't know if the word is redundant or superfluous. Because now you still could have a reasonably high-performing place that is not running away from the identification of needing to do something about a target population. But the rhetoric of the big system—I am either in or out—it is not effective.

Mr. KILDEE. Thank you, Commissioner. Thank you very much.

Thank you, Mr. Chairman.

Chairman MILLER. Thank you.

Mr. Castle?

Mr. CASTLE. Thank you, Mr. Chairman.

And let me thank the panel of witnesses who were exceptional. I started this week giving a speech to our district superintendents on the growth model. And then I have listened to you. And I have decided now I knew a lot less about it than I thought I did going into it. So you have opened up the book for study, I think, here.

Let me ask you, Dr. Doran, a question on something a little bit unrelated in your written testimony, which I am looking at now. You indicated in the discussion on NAPE, "It cannot be used to measure AYP." I agree with that. "But it could be used to inform how state performance standards are set," et cetera, and, "recommend that the committee support a research agenda that would investigate and report how best to use links between NAEP and the various state assessment programs across the country." And I agree with that, too.

And I have seen the charts that have shown how states are achieving on their own assessments versus how they do on the NAEP test, the National Assessment for Education Progress Test. And I would assume the state assessment would include standards, too. I mean, to me they are perhaps—I am not saying anyone is cheating. But obviously, some states are setting a lot higher standards than others.

And that concerns me. I am not sure that is what the purpose of all this is. But I just wonder if you wanted to expand on that a little bit in terms of your thinking. I understand your conclusion is we need to study it further.

Mr. DORAN. I would be happy to. It is true. We know that there is a lot of variability in at least the two things that you mentioned. We know the difficulty of the assessments vary across states. And we also know the difficulty and the breadth of the content standards vary across the states.

And there is some research. It is not comprehensive. But there is some research that has done exactly what you mentioned. We have seen how state tests can be used to match up to NAEP. And we can compare how state performance compares to NAEP. I think we need to extend that, and that is why I am recommending that. I would like to see that happen a bit more comprehensively.

I think this is important for a number of reasons. One, I am not sure that there is a great deal of understanding of exactly what is happening in or why there is this great variability across states. And I think we need to open the door to start having that conversation about if there is variability, what is the cause of that variability, and are some states, in fact, doing things that other states should be doing.

So I think having a policy that would help illustrate the comparability of standards and assessments across states would then lead us down the path of a better understanding about what some states are doing that may, in fact, should be replicated in other locations. Why do I think that is important? Well, we know that some students start high school in one state and they move into another state. And they may have a difficult time catching up. Or they may be advanced, and they may be bored.

That some students graduate high school in, say, Arizona and they may move and attend college in California or obtain work in California. But the proficiency definitions in Arizona and California may be very disparate.

So we in many respects don't have a really strong system of coherence. And I know why. Because we have—someone mentioned 50 or 52 different experiments happening with the district in Puerto Rico. So I would recommend this because I think, a, we need an illustration of what is happening in terms of comparability. And, two, I think that would lead us down the road of a better understanding of why there are variances.

Mr. CASTLE. Thank you.

Let me jump subjects here and to Secretary Woodruff and Dr. Dougherty, getting back to the growth model.

Secretary Woodruff, you mentioned that Delaware has been using longitudinal data systems that track individual student progress since 1984. And my impression is from your testimony

and from what I know that indeed Delaware was more advanced in that area than perhaps some other states had been.

Dr. Dougherty, I think you indicated that 27 states could do growth now and 40 in several years. Is that a correct statement? Well, let me ask the question. And that is the whole growth business is a little more complicated than I had thought, I am learning. And my concern is—and I think it is an important part of our discussion on the reiteration of No Child Left Behind perhaps this year.

But my concern is the ability of the states to do it. We have had a lot of complaints about the cost of No Child Left Behind, et cetera. And I don't want to overburden. On the other hand, I would like to do something which is positive. I am just curious as to where we are vis-a-vis the states and how simplistic this would be for them to do or how complicated it would be for them to do it. If you all could share your thoughts on that.

Ms. WOODRUFF. Well, I think Dr. Dougherty certainly is much more the expert on the lay of the land, if you will, across states and where different states are. But I know that in our conversations at CCSSO that as states are putting the data systems in place and learning more about assessment systems and how growth can work, there is a desire among my colleagues for this kind of accountability model to be used because we feel that it really can help us, quite frankly, incentivize our schools and people within our schools more than the status model alone.

Mr. DOUGHERTY. Yes, I would say that basically the data nerds in the state agencies have been wanting longitudinal data systems for years. And they never got the leverage until No Child Left Behind came along and you started to talk about, "Well, you have got to desegregate kids by ethnicity," and so forth and so on. And then how do you keep track of which kid belongs in which group with kids bubbling in every year? That is going to create errors and so forth.

And so, you basically—one of the biggest positive consequences of No Child Left Behind is just the better development of data systems and the greater use of data for school improvement, system evaluation, and so forth. There has been tremendous progress. My organization was originally a small non-profit called Just for the Kids. And we started out in 2000 surveying the states to see who could do longitudinal data pictures involving student growth, tracking, who has been enrolled in the school for how long.

And Tom Luce, who founded our organization, said, you know, find me 15 states that can do this. Well, we found about five. So now it is a lot more than 15, so we are making tremendous progress in this area. The recognition that it is valuable, that it is not only valuable for accountability, but you can then put information in the hands of educators.

I have not only got my kid, but my kid comes in this fall and I have got an academic history on the kid going back. So if he doesn't understand multiplication, maybe he didn't learn place value last year. Understanding that building these data systems is valuable, both for evaluation, accountability, and school improvement and the teacher and principal and district level.

Mr. CASTLE. Thank you.

Thank you, Mr. Chairman.

Chairman MILLER. Ms. Hirono?

Ms. HIRONO. Thank you, Mr. Chairman.

I think that NCLB should allow for multiple assessments because what we have now is just not fixable enough as a really helpful way to measure student progress. And right now the Department of Education is approving growth models on a pilot basis. And they are limiting this to only 10 states.

I note, Dr. Dougherty, that in your testimony that 27 states are pretty much ready to go with a growth model and that the NCLB right now does not contemplate that by statute.

So yes or no would be good, for all of you, if we should amend NCLB to allow for more flexibility to allow the states right now to propose a growth model as an assessment measure. Can we just go down the line?

Mr. OLSON. Yes.

Ms. WOODRUFF. Yes.

Mr. DOUGHERTY. Yes.

Mr. MCWALTERS. Yes.

Mr. DORAN. Yes.

Ms. HIRONO. Thank you.

Thank you, Mr. Chairman.

Chairman MILLER. I am impressed. Thank you, Ms. Hirono.

We will go back to Mr. Boustany.

Mr. BOUSTANY. Thank you, Mr. Chairman.

Given that math and science have been—there is a strong consensus that these areas of education are critical for our national competitiveness vis-a-vis China and other countries in a global economy.

For those of you who have looked at the longitudinal tracking, are there clear differences with regard to math versus language arts when you look at the tracking system? And is it easier to implement longitudinal tracking with math education than with language arts?

Mr. DORAN. I have done a bit of research on this actually. It is a tough question to answer. It is a good question. And we think about this question quite a bit actually.

In the growth modeling world, and in a slight variation from the kinds of growth models that we are talking about today, something called value added models, we tend to be able to pick up what statisticians call a bit more signal, that is, we can minimize statistical noise, with math. We don't know exactly why.

Some hypothesize that math tends to be a little bit more of a linear kind of an instructional subject as opposed to reading, which one may or may not—and there are arguments on the other side of that, that they say math isn't as linear. But from a statistical perspective in some of the research that I have done with value added models, which are slightly different than growth models that we are talking about here today, we are able to at least pick up a bit more sensitivity on what is happening within the school in the subject area of math.

We still do very good work with—or we think we can still do very good work statistically with reading scores. But the sensitivity in terms of how much we can capture for whatever reason isn't as

good in reading as it is in math. It is still good, and I don't want to undermine that it is not. But we can pick up better patterns of what is happening in schools and minimize statistical noise with math when compared to reading.

Mr. MCWALTERS. I think the issue of math and reading comprehension and communication are the central elements. My experience in this is that we have to delve deeper into what reading comprehension means. And testing has its limits there.

But in the industry that I represent, people understand teaching reading. And yet they stop teaching it developmentally by the 4th grade, which is why you have so many kids who can't answer comprehension questions when they get into high school. And math is too often defined as operations as opposed to problem solving.

And my experience is that once you are in high school, a student who can't solve the math problem probably isn't reading and comprehending what you are even asking them to do. If you can reduce it to an operation, they tend to be able to do it.

I have kids who can pass an algebra test if it is done as algebra problems. If you take the numbers off the page and put it as a problem to be configured and then solved, they can't do it.

Mr. BOUSTANY. So there is a strong linkage between language skills and math solving ability.

Mr. MCWALTERS. At the higher up that you tend to go.

Mr. BOUSTANY. The higher up you go?

Mr. MCWALTERS. Absolutely. Problem solving——

Mr. BOUSTANY. So it is critical that if we are going to use longitudinal tracking as a tool, you wouldn't want to separate out the two. You would want to track both areas longitudinally?

Mr. MCWALTERS. Yes, absolutely.

Mr. BOUSTANY. Yes, yes.

Also, Dr. Doran, I was very pleased to hear your commentary on the variability of NAEP and many of the state assessments. And this seems to be something that has been unmasked clearly since No Child Left Behind has been in play. And I agree. I think it is an area clearly that needs to be researched more thoroughly. And so, I thank you for bringing up that point.

I see my time is running out.

Mr. Chairman, I will yield back.

Chairman MILLER. Ms. Davis?

Mrs. DAVIS OF CALIFORNIA. Thank you, Mr. Chairman.

Thank you all for being here. I really appreciate your expertise on this.

Commissioner McWalters, you mentioned one of the problems that we have certainly seen in the San Diego area where we had a school meeting AYP on one of 30—only missing it on one of 38 requirements.

In your research, if we were to address the specific shortfalls for a school and just look at that element—and in many cases, it is in special education or perhaps even in English language learners. Does that actually cover the needs for that school? Or how do you think we should best address that?

Mr. MCWALTERS. Now you are into context. And take this as an experienced practitioner, but it is not definitive. I can imagine a place where you expose the one indicator and the people are as

upset at the school level as we would be. It is almost like when we finally got decent information that has surfaced, they are willing to step toward the problem.

There is another school where that one indicator—those kids become the problem. They will do everything they can to find a way around the kid. Those are two different contexts. One of them you want to hang. And the other one you want to work with.

Now, however we term that, this is that issue of is everything too blunt. Assuming that you have taught us the lesson that we are accountable and that we have got to be transparent, this tension between state, district, and school has to come to a new level of maturity where I am holding the right issues and people accountable for the right attitudes and intervention strategies. That is the best answer I can give you.

Mrs. DAVIS OF CALIFORNIA. Anybody else want to address that? Okay. It is obviously a difficulty in the community. It is a huge difficulty for schools. And I was just curious to see how many people have——

Mr. MCWALTERS. But I want to say again. I have communities also that want those kids then to be isolated. That is the good part of NCLB is that these are all our kids. And to the extent we are on task to solve that problem, I need to be an incenter, a rewarder, and a partner. If you are avoiding those kids at the community or district level, I need to be the hammer.

Mrs. DAVIS OF CALIFORNIA. Yes. And perhaps this is an expansion on that a little bit because we know that there are certain subgroups that are more likely in some school districts to not meet the requirements. And there is this tension, as you say, with identifying certain sub-groups. Is there a growth model, though, for those sub-groups that might be more pertinent really within the context?

For example, in English language learners, you may have a classroom where you are moving the kids out of that classroom. The fact that that classroom isn't showing improvement isn't because the kids aren't improving. It is because the kids who did improve moved out of the classroom.

Mr. MCWALTERS. Yes.

Mrs. DAVIS OF CALIFORNIA. How can we best demonstrate this concern? And in many ways, is there just a downside to the growth model as well?

Ms. WOODRUFF. If I could respond, actually when we designed ours—and our growth model will give schools and teachers within schools more specific information about individual students. And in particular, one of our directors for special education in one of our local school districts is really intent on this particular model because we will be able to see if a student is making that kind of progress and then they can then examine what needs to be done for that particular child.

One of the other things—so I think the growth model does really incentivize and provide additional information, more in-depth information for schools and districts to be able to act. And I think that is important.

The other thing that we are finding as we look at the issues around English language learners and special education is—and we have done a lot of work to try to build the capacity of local districts.

That even though you may have schools within a district that are kind of going up and down, that it is a district level issue that needs to be dealt with. And we need to help them intervene across the district, not just in individual schools so that you can stop some of that fluctuation.

Mrs. DAVIS OF CALIFORNIA. And, Mr. Olson, perhaps if you want to come up really quickly. We are running out of time. But I just wonder is there good cooperation between states with this data sharing and in developing the longitudinal work that is being done? Do you see some really good examples that we could look at?

And, Mr. Olson, did you want to comment on the last one real quickly? Mr. Olson, I am sorry. Did you want to comment on that last comment?

Mr. OLSON. I wanted to comment on your earlier question. Given our work, students typically will take a test two or three, four times during a year giving accurate information on the growth measured. When a student moves from one classroom to another, we have the data that follows the child.

The other interesting thing is that with the kind of quality that we bring, the information we bring, we can begin studying the effects of moving a child from classroom to classroom.

So it may or may not be—you know, if the student achieves somewhat less or less growth, it may not be the teacher. It may be the fact that the child was moved from class to class or from school to school. But the quality of data allows us to begin understanding issues like that within the school system.

Mrs. DAVIS OF CALIFORNIA. Thank you.

Mr. DOUGHERTY. In answer to your question about cooperation across state, states are ravenous for information about how other states are doing it. One of our most popular things that we have got in the data quality campaign has been to do a lessons learned series where we have gone out and done detailed site visits in specific states and said, "How did they go through the process?"

This is stuff that is difficult to record in a survey, so it is their nuanced experiences. This has been in very high demand in other states.

Chairman MILLER. Thank you.

Mr. Souder?

Mr. SOUDER. I have a couple of questions. But I didn't hear a clear answer to Mr. Castle's earlier question.

And maybe, Mr. Olson, you could take first crack. How much roughly does a growth model increase the costs?

Mr. OLSON. We don't have all the data about the costs for each school district, I mean, each state. But our measures in all likelihood could be put in place for a state at a cost real similar to what a state has gained for measure for one time a year under the current model.

Mr. SOUDER. Thank you.

We spend—and schools spend even more—millions on IDEA and developing individual education plans that supposedly are advancing those special needs students at the best rate possible.

Does the growth model accommodate that? Is anybody talking about how to integrate what we are spending with the right hand into the left-hand measurements?

Mr. OLSON. Well, I would just make one comment. And earlier my remarks focused on two things. One is measuring growth. And two, measuring individual children accurately enough to measure growth. With the computerized adaptive measure, which we use—and there will be other methodologies.

But when you are measuring children accurately, we can measure academic growth of children about 98 percent of the children within the normal population. Which means we are measuring accurately academic growth of most of the special education population as well as most of our most talented children. So a real good accurate measure plus growth is applicable to those programs.

Mr. MCWALTERS. Can I comment on that?

Mr. SOUDER. Yes. I would also like to see how that is integrated in, then, to the individual education plans and whether these two things are actually linked at all in the real world.

Mr. MCWALTERS. I think that is the right question that has a complicated answer. One is it is No Child Left Behind that finally got on the table that other than for a small number of students we should have the same standards for all kids.

I am the parent of a special needs student who finally graduated from college sum cum laude in math who could not possibly pass any of these tests as a 4th-, 5th-or 6th-grader. So the issues of adaptations are very real. But the issues of common standard expectations need to be pounded on. That is the right place to be.

Now, having said that, the instrumentation for changing expectations and changing classroom practice is we have so far to go that the AYP exercise right now is almost likely to pick up all of the common cultural heritage that we didn't expect these kids to do anything. So the intervention strategies now have to be comprehensive. They have to be intensive.

But we have to be realistic about where we are starting. And I do think working that back into the individual improvement plan strategy and logic is a pretty powerful institutional problem that we are facing. And that is the only way you are going to bring assessment and IEPs into kind of a common mission.

Mr. SOUDER. Because most of the schools in my district who are failing in the standards are either special needs, or the second is ESL. Because clearly, you can almost tell uniformly ESL mix even in Indiana. It varies even in a district. Some buildings will have 80 percent, and others will have a small percent.

Some, very few, that are failing—I mean, a school can waiver a certain amount. But most of the schools that are having problems are way over the amount that they are allowed for a waiver. What we have been talking about today—how does that integrate with the English as a second language?

Mr. DOUGHERTY. I want to mention that an ESL kid is particularly likely—one who is just learning English—is particularly likely to be very far below proficiency on an English language test, since he can't read the test, at the beginning and then is likely to make very rapid progress. So you should note a rapid growth trajectory for such a student.

Some states, of course, do have tests that measure the kid's progress in learning English. And school systems use those tests as part of their diagnostic understanding of why the kid isn't pro-

ficient on the English language test. It is because they are not proficient on the test of English proficiency. California is a great example of a state that has really been conscientious in developing a test that tracks kids' progress in learning English.

Mr. MCWALTERS. The huge difference is in grade spans. If you are somebody coming in here at 2nd grade coming from some schooling, first of all, by age you are developmentally more likely to respond to whatever the treatment is. If you come into the 10th grade with no schooling, that is a different treatment.

I think we shouldn't confuse measuring the measurement of capacity or fluidity in a language with the other issues behind the individual child. This is much more about program treatment, the integrity of good program treatment in the ELL world while we are figuring out the different ways to measure what it is, language capacity or language fluidity, either readiness or in English. I think those are—we have to separate those issues and go after the integrity of program treatment because there is tremendous variability on these children.

Mr. SOUDER. I had a young student from, I believe it was, Southside High School in Fort Wayne, Indiana, who had come in from Somalia where we have a lot of refugees coming in from Eastern Africa. And he said first off, he was given the test 30 days after he arrived and spoke no English. And then even after he learned English, they had never taught math in Somalia. So even after he became proficient in English for his grade level, he was substantially behind.

Mr. MCWALTERS. Right, right.

Mr. SOUDER. These nuances are just devastating to some of the morale to the teachers. I mean, I want accountability. But it is devastating to the morale of the school and the teachers when they are being measured and told they are failing based on those kinds of standards.

Mr. MCWALTERS. Right. But we have also many students in our country that are American, as in born here. And they are growing up in second language homes and neighborhoods. And they are not doing well in our tests, either. That isn't about measurement. That is about program quality. And this is about the intervention strategies.

So I think the whole ESL question is the right question on the table. And the issues about language facility in their own language and in our language—all of that I think we have measures of that. But how you fold that into an accountability system and a program intervention question—it is not solved in the timelines that are in the NCLB exercise.

Chairman MILLER. Mr. Hinojosa?

Mr. HINOJOSA. Thank you, Mr. Chairman.

I thank the panelists for coming into visit with us today and telling us what your thoughts are on No Child Left Behind.

My first question I am going to direct to Peter McWalters and to Valerie Woodruff. No Child Left Behind already requires a growth model in one area. And that is for limited English proficient students. States are required to have benchmarks for English language proficiency that are aligned to the state's academic content standards.

They are also required to annually measure students' progress toward proficiency. Share with us what steps your state has taken in implementing these provisions and how your experience with LEP students might inform our approach to growth models of accountability.

Mr. MCWALTERS. We are part of a national consortium to try to come up with both assessments and treatment. And as I said just a minute ago, so far to protect the interests of everybody, we have all of them tested in state testing, and we report them as disaggregation so that it is still currently transparent. It is only through that exercise that I think that I have not got the other layers of information, which is I have some students where that is a good measure of the system's failure to treat them.

I have other students that shouldn't be taking that test. And it is almost a keen sense of the obvious when you see that. So I am trying to help people understand we have got to figure out the measurement instrument, which is necessary. But I think we also have to know that in some cases it isn't about the measurement.

It is about the program that that child is in and either the integrity of its delivery or the fact that he shouldn't or she shouldn't be in that program. And I am trying to play that out right now both ways. But I am using straightforward state assessments to do it. And that is why that cohort isn't moving because many of them will not show significant enough improvement fast enough to get that program—or those kids off that list.

Mr. HINOJOSA. Valerie, what is your state doing?

Ms. WOODRUFF. We certainly are measuring the students' proficiency in English. And I would agree with Peter. There are a number of children who, because of their varying circumstances, 12 years old, no schooling in their native language coming to us and then we are trying to catch them up, who really should not be participating in the state assessments. They do to the extent that they can. And that certainly tells us where they are.

But we really need to be held accountable, in my mind, for particularly those older children and whether or not they are meeting proficiency in English first and then become part of our state assessment system. So, we actually implemented a test of English proficiency before No Child Left Behind and required our districts to track them. Also, once those children become proficient, we require our districts to continue to monitor how those children are doing. And if they begin to falter, then to intervene and provide additional support.

So that has been something that has been kind of on the books and in practice in our state for a while. But we continue to be concerned with the frustration level of the children who are required to take an assessment that they cannot begin to understand and much less, be proficient on.

Mr. HINOJOSA. The next question I want to direct to Harold Doran and to Allan Olson. No Child Left Behind's accountability measures are least effective in high schools and is proven by how we are competing internationally. Our high schools are way, way down on the list as compared to China and Singapore and all those others.

What are your recommendations for meaningful accountability at the high-school level that would include multiple measures, readiness for both secondary opportunities, and real progress on improving graduation rates?

Mr. DORAN. I have a couple thoughts. And I was wondering actually if that question would come up in today's conversation. Bill Gates gave testimony here a week or two ago, and this issue was highlighted. And there have been some recent studies that I think have been illustrative of exactly what you are talking about.

I think there are a couple of things that I have learned by looking at the literature recently that have evaluated state assessment systems that tell an interesting story. I may get my numbers slightly wrong, but I think the number is something like this coming from Project Achieve and some studies they have recently done.

I think it is eight states have aligned their graduation requirements with expectations for post-secondary education or the workforce. Twenty-six states have their assessments, high-school assessments in place that only measure skills that measure 8th-, 9th-and 10th-grade skills. And those don't necessarily translate into skills that would guarantee that students are successful post-high school.

There is an interesting model that 11 states have recently bought into. And they have formed a consortium around an Algebra II test. And the idea here is that when students demonstrate competency in Algebra II that that guarantees—or at least that gives them a higher probability that they will be successful post-high school. And in some of those 11 states that test will be a graduation requirement. In some other states, it will not.

But I think one of the things that we can do from a policy perspective is ask the following question. What do we want for our children, and how do we know we are getting it? And so, one of the things that we ought to—that we want for our children is success post-high school. We need to operationalize and define what that means.

Eleven states—there are more probably doing it, but I can cite the example of 11 states. They have said we value Algebra II. How do we know we are getting there? Well, we are going to measure their progress on that core content area because 11 states—we believe that should students demonstrate competency in that particular content area, they are likely to be successful in high school.

So I think we can start with something simple. Ask the question what do we want for our children. We want success in post-secondary education. And what does that mean? And then implement systems that measure that.

Mr. HINOJOSA. Thank you.

Chairman MILLER. Thank you.

The gentleman's time is expired.

Mr. Heller?

Mr. HELLER. Thank you, Mr. Chairman.

Just a couple of questions here. And I appreciate the panel being here. I really do appreciate your input. You guys are the experts. I am not. My wife is a school teacher, so every once in a while she does chew on my ears a little bit, especially on this particular topic.

And one of the issues probably reflects what the ranking member was saying and her concern about the amount of time you spend

testing children as opposed to the amount of time you actually teach children. And it flows over.

For example, I represent Northern Nevada. And the elementary school that my children go to, because of the amount of teaching—excuse me, the amount of testing that goes on, they have dropped certain curriculum. For example, they don't teach history any more in the elementary school level because it is not tested under NCLB.

They have dropped geography. They have dropped social studies. And that doesn't include other curriculum or activities like the music programs. They are dropping all these programs because they are so concerned about these core issues that need to be taught and tested that they don't have time to teach others.

And I was wondering perhaps, Ms. Woodruff, if you could comment on that.

Ms. WOODRUFF. I would be happy to.

Mr. HELLER. Thank you.

Ms. WOODRUFF. I think those schools are wrong in their dropping of those other curricular areas. Interestingly enough, in Delaware we assess both science and social studies and have been doing so at the elementary-, middle-and high-school level since 2000. And we will continue to assess social studies as well.

The other piece of that is that I think that when schools begin to eliminate the social sciences, when they begin to eliminate the arts programs, they are failing to see that there is another context within which children can learn reading and mathematics.

Mr. HELLER. I agree.

Ms. WOODRUFF. When children see the relationship to other kinds of—to the rest of their lives and to other kinds of learning, they are much more likely to be successful than if they are being constantly bombarded with only two or three particular subject areas. There are a number of research studies about the arts and so forth. I just think that it is something that I am very happy to tell you the schools in Delaware have not done and that we encourage them to understand how those linkages can be made.

Mr. HELLER. And I agree with you because I think that is an imperative part of a child's education, are some of these social skills that they learn in this process.

Ms. WOODRUFF. Right.

Mr. HELLER. And I guess my concern, Mr. Chairman, is that we are limiting the curriculum of these children or are careful not to limit the curriculum of these students because I think music programs do offer value. I think history offers a lot of value as does geography and other social studies areas.

Ms. WOODRUFF. If I could comment further, one of the things that we have deliberately done is we have standards in about 17 different areas, including career and technical education. And we have done crosswalks, if you will, between standards in one area and in another so that the people see the relationship.

We are also in the process of developing a statewide recommended curriculum with model units. And many of those units are integrated so that the teachers have something that they can use. And then they have embedded assessments that are directly related to the instruction that then just flow out of the whole teaching and learning process and are not seen as some stand-

alone test that they don't feel has any sense and context of the school itself and of their ongoing work. So it is really an exciting opportunity for us.

Mr. HELLER. Okay.

Ms. WOODRUFF. And our teachers are helping us build it and are embracing it.

Mr. MCWALTERS. I think this is a wonderful opportunity to get out of the silos by the cross-mapping. I am assuming most people would still want reading and math assessed because they are so central. The idea that they are displacing something or teaching to the tests as in drill and kill obviously is the wrong place to be.

But when you start helping people map across the subjects, then all the activity, the actual hands-on applied learning exhibitions become instrumental in improving those two scores. That is one of the only ways we are going to change the structure of schooling. Otherwise you are going to end up with more separation, more discrete testing. And it will still be factual recall rather than application.

Mr. HELLER. Thank you, Mr. Chairman.

Chairman MILLER. Thank you very much. I guess I would argue that when schools start to implode on narrowing the curriculum it may be one of the first indicators of the lack of capacity, that you really are now watching an institution that is atrophying to such an extent and lost an understanding of what a learning environment is.

I mean, I have been involved with a number of schools all across this country that have now taken music and made it an absolute gateway to mathematics and the understanding of mathematics. And I mean, it is replicated time and again in so many areas that that might be a red flag that you would not want to ignore in terms of the talent of that group of teachers and administrators.

Next is Mr. Courtney.

Mr. COURTNEY. Thank you, Mr. Chairman.

And I was out of the room for a minute, and you may have covered some areas while I was gone. But I am from the state that is suing the federal government over No Child Left Behind, which is the way I was introducing myself at a lot of workshops for freshman members. And to be honest with you, it was actually a fairly——

Chairman MILLER. That would be Connecticut, right?

Mr. COURTNEY. That is correct. Sorry.

And, you know, listening to the presentation, which obviously all of you put a lot of thought into what is the goals—which I think everybody agrees on. But I have to say there really—at least in the state that has distinguished itself in terms of the hostility and adversarial relationship with this program—it is a very popular thing that the attorney general is doing.

He is the kind of guy who sues everybody, pharmaceutical companies, banks, insurance companies. He has said that there has been no action of his office that has ever garnered the kind of public response as his decision to challenge NCLB.

And, Mr. McWalters, who is a close neighbor of my district——

Mr. MCWALTERS. I am?

Mr. COURTNEY. You sort of started to get into whether or not there is sort of a redundancy factor about what we are sort of learning from tests. And, you know, what I see in Connecticut is that when the test results come back in, the schools that are not succeeding are Title I schools. And, I mean, it doesn't take a rocket scientist to figure out that Greenwich High School is not going to have any problems succeeding. Whereas New London or Willington or Hartford or Bridgeport or New Haven are going to—I mean, and at some point people really question about, you know, why is this effort and expense worth it?

Because it is almost common sense that tells you what the results are, which is we know where the problems are. It is poor school districts who, by the way, are the ones who have been getting shortchanged on Title I funding over the last couple of years. I mean, it is almost perverse to see the cuts that these districts are having to absorb over the last few years in terms of resources.

At the same time, the government is identifying them as not succeeding. So, you know, I guess the question is is there a way to do this a little more intelligently without sort of, again, really damaging the public's belief and credibility in a process that they see as the tail wagging the dog.

You have to go back now to the beginning because the issue—some of us experienced this between the law as passed by a bipartisan Congress with an executive branch that was drawing a new line in the sand for accountability and transparency. That is good. Disaggregation, good.

Many of us were in states that had systems that pre-dated that. Valerie spoke to it. I spoke to it. I can clearly remember sitting with the department going, "Wow, what an opportunity."

If you came in and assessed where each state was in terms of its own integrity to do the right thing—as in we had just got a law. I was into disaggregation. I was into the beginning of intervention. It didn't line up perfectly, but I was there.

Instead of leveraging me forward, I spent 18 months regrouping. That was a mistake. But I write it off because I think the impatience of Congress from a nation at risk to goals 2000 was such that you didn't want to hear it anymore. Connecticut was a perfect example, a high-performing state with some of the biggest gaps, some of the most urban concentrations.

The way to call that question between the law and the department to focus in on what needed to be called apparently didn't happen. I was one of those states that said I don't need more state testing to know the sick place. But I have learned to appreciate grade-level testing as an instrument of improvement at the school and district level. I couldn't have untangled that 5 years ago.

But I think we are all saying whatever lessons we needed to learn about accountability and capacity and transparency—if it hasn't been learned, then you need to authorize your department to go after that state. But for states that have stepped toward this and they are trying to sort out state needs from district needs to school needs to growth, individual, instructional needs, we have got to get that sophisticated pretty quickly.

Ms. WOODRUFF. Well, our experience has been that many of our Title I schools are some of our highest-performing schools. And we,

for a very small state, are thrilled that we have had a number of national blue ribbon schools. High-poverty schools with high-risk populations, including English language learners, particularly at the elementary level who are doing incredibly well.

I think that where we are seeing No Child Left Behind really shining the light on places that may have a somewhat homogenous population and smaller numbers of the sub-groups and shining the light on those places and saying you are not doing what you need to be doing for all children has been helpful in many ways. And I think that Title I schools for many years have received a great deal of money.

I am not happy with the way Title I schools have to hold back certain amounts of money in case of choice and in case of supplemental educational services that should be, in my mind, going to programs and to children rather than being held back for some of those reasons. But our experience with Title I, non-Title I has been a little different than what you described.

Chairman MILLER. The gentleman's time is up.

Mr. Fortuno?

Mr. FORTUNO. Thank you, Mr. Chairman. First of all, I want to thank you for holding this hearing today, and the ranking member as well.

And thank you all for being here. I am sorry I had to step out for a while. But as I was following everything that you have said—and I was here through all of your presentations today—it is clear that there are different states at different levels of achievement. Some states have really benefited from this process. And actually all their resources have been focused in trying to do what needs to happen.

The other states like Connecticut—I would love to share some thoughts with you afterwards, if we may—certainly are not as happy. In my case in Puerto Rico actually, the latest was that the AYP measurements or actually requirements are not being met, and Puerto Rico was just fined a couple of weeks ago on this.

And actually, if I may, Mr. Chairman, I would love to introduce that letter just to show that indeed there are different jurisdictions at different levels of achievement. So if you don't have any problems with that, I would love to introduce in the record the letter from the Department of Education.

And I am just wondering when you have this disparity—and I am asking everyone—how do you handle—from here, how do you handle that disparity. We want some levels of measurement. There are some states achieving—or actually some districts that are at a very high level of achievement. And there are other places like my district where that is not happening, clearly not happening.

So I would love to hear your insights as to what you recommend we do from this end to try to do something that fits everyone. But actually it is impossible to fit everyone.

Mr. MCWALTERS. I want to step right up on that one. I think the law was trying to protect the rights of children to get—to access to a quality education. And thank God, it holds states accountable for that. That is the right place to be.

But having said, I am the smallest geographic state, but I am the second most densely populated state in the union. We have about

the same number of population. We were comparing demographics earlier.

Every one of these places has a very different issue. And I would still submit that this law does not address what the original Title I law was trying to do, which was become an issue—I am going to call it the urban agenda—in concentration and size. They make a difference.

If you are dealing with New York City, Chicago, Philadelphia or Los Angeles or Providence, which is a small big city, the issues you are dealing with to get the individual school and student access in quality instruction is complicated by distance, size, and density.

I think the law hints at that, but I don't think there is enough understanding that for me to get a child in New York City access and performance to standard is to deal with all of the issues from state house through district to community. And it is somewhere in the differentiated instruction. It is the same in Connecticut.

Connecticut's issue is basically urban. Now, I don't know whether the state had an urban agenda. As a superintendent, I don't know many states that did, at least not really. Because if it is an urban agenda, it is more complicated than simply school improvement, as necessary as the school improvement infrastructure is.

Mr. FORTUNO. Anybody else have a comment? The weather is great in Puerto Rico this time of the year. And if anyone wants to come down, I guarantee good weather.

Ms. WOODRUFF. I think your point that the law—although the goals of the law are certainly well-intended—that just as we know that every school has its own unique needs and issues, every state is in a different place. And I think that Peter mentioned earlier that, you know, if you have a state that has put systems in place and is moving forward and getting the agenda taken care of, then we ought to be allowed to do that and to be given some freedom and flexibility in order to do it.

And those folks who are seeking to improve, such as Puerto Rico, need to be given technical assistance and support. It is part of what we keep talking about in terms of a federal, state partnership. And a partnership is you shake hands, and you figure out where you are going, and you help each other get there. It is not a one-size-fits-all and everybody lines up and you are either yes or no and put in a box. That is part of what has been frustrating.

I believe that when No Child Left Behind was passed that Delaware could have made minor changes in our existing law, and we could have been much further along today than we are because we had to sit back and regroup. And I was told point blank by counsel at the Department of Education that our law was too restrictive and it needed to be changed. Our law was changed. And we are now in compliance with No Child Left Behind. We would be in a very different place today, I believe.

Mr. FORTUNO. Thank you.

Mr. OLSON. I would like to make an observation. The law when it was passed was passed with the intent to use an accountability process to help schools and states get better. Well-intended, well-conceived. But a message within the panel today is that there is also a need for Congress to reflect on how might that law be more

helpful to the processes of improving learning and instruction and school organization and things like that.

I think if you reflect on that question, given the resources that are being put in and the issues that you have within—the consequential issues you have within the law—and I am not suggesting—that reflection—I am not suggesting walking away from any of the requirements. But how might the law change in small ways to make it easier for schools to put the energy into constant improvement over time?

Mr. SARBANES. Actually I want to pick up on that idea—excuse me—and talk about and ask you a couple questions about this relationship between resources and an accountability framework, which is usually put in the context of well, we have the accountability and we just need to get more resources in there behind what people are doing so that they can actually achieve the goal.

But what I am interested in having you speak to is whether, for example, you think a growth model that has been discussed in contrast to this status model, whether that can actually result in more efficient use of resources.

I mean, I had the opportunity to be part of rolling No Child Left Behind out in the state and the district and the district of schools and now within schools. And as you know, the current system is such that when you don't meet AYP particularly for certain periods of time, it triggers all kinds of technical assistance and other resources and requirements on the system and then on schools in terms of developing school improvement plans and restructuring plans and all this other kind of stuff.

So that is an obvious place where if a growth model brought more flexibility into the system and the accountability system you might not start a school or a system or a series of schools jumping through those hoops that then generate a lot of resources as quickly. So you could speak to that.

But then the other question is just in the delivery of resources do you think a growth model is going to encourage the resources to be directed better than they are being directed now?

So I would love to have you all react to that question.

Mr. OLSON. I would like to make a brief comment.

Mr. SARBANES. Yes.

Mr. OLSON. And I will go back. There have been a number of observations that schools, say, were tested too much. What we find in our work is that once people are administering tests that are useful, helpful and drive improvement of their decision making, all of a sudden they think of testing as being desirable. So a lot of it has to do with the utility and the accuracy and the helpfulness, if you will, of the measure.

So if you go to a measure that is more accurate than that which is commonly used, right away you, if you will, free up resources and you change the resource allocation. You change the energy. You change the decision making.

If we have improved information about growth and about growth of individual children, we will then also know more about the factors of the resources and which are making a difference. And so, we can make better decisions which to use, which to modify and

how to use them. The growth measure, a good, accurate growth measure will, in fact, influence resource allocation over time.

Mr. SARBANES. All right. Thank you.

Ms. WOODRUFF. I would agree. We know that with the implementation of our growth model—and we have already done a few test cases and given information to some of our schools—that they are able then to hone in on specific children a little differently. And to go to the gentleman's question a little while ago about the use of IDEA funds and so forth, we foresee—and I think this will hold true, continue to hold true—that the allocation of resources toward specific needs of not only groups of children, but individual children will be more targeted.

I think then that as that happens, we ought to be given some flexibility in how to utilize those funds a little differently than perhaps we are required to do today. And I think that the whole issue of resources needs to be examined in terms of the efficiency with which schools and districts are using the resources at hand. Not to say that we couldn't use additional resources, but the examination of efficiencies is always important.

Mr. SARBANES. Right.

Anybody else?

Mr. DORAN. Yes. The interesting thing about the growth model is it can tell a very different story about a school. And this would very directly interpret or suggest how we would do resource allocation. For example, the current system says you don't cross the threshold, you might be low-performing. If you are above the threshold, you are making AYP or some might perceive that as being high-performing. But it may, in fact, be the opposite story that we want to be told.

In fact, we may have students who are very high-performing but they are dropping in their performance. The school is not actually doing a good job with those kids, but they are staying above the proficiency bar. On the other hand, we might have a school that is doing a very remarkable job with low-performing kids. They are not getting them to cross that proficiency bar just yet. They will, but they didn't just yet.

Now, in fact, it is the school that appears to be high-performing under the status model that actually needs resources targeted to it. And it is the other school that is doing a remarkable job with its struggling kids that appears to be doing okay. We do the opposite right now, not in all cases, but in many cases. And so, that would have a direct relationship.

You know, we see this happen in other fields. And I talk to educators often about how to make good use of data and definitely explore different statistical methods. We recently saw this happen in a book that illustrated how, when you have the autonomy to look at statistics and data and mine your data, how you can figure out how to build better teams.

I think we would see the same kind of thing happen in education. The autonomy to use better and newer statistical methods will allow for us to figure out how to build better schools.

Mr. SARBANES. Thank you.

Chairman MILLER. The gentleman's time is expired.

I would just comment on the gentleman's question because I think it is a critical question, one as to flexibility of the use of resources and how you use the data. But, you know, in every other segment of the economy, people have been plowing the resources to developing data so that they can make smarter use of human capital or capital budgets and all of the rest of this.

I mean, all across the board that is the competition that is taking place within the economy. And this is one of the—this and health care are sort of the last areas to decide that data can really improve the deployment of resources and the efficiency of those resources.

Mr. Keller?

Mr. KELLER. Thank you, Mr. Chairman.

Mr. Chairman, I think it is critical that we get this bill in the strike zone or it is going to be in trouble.

On the right, conservatives don't like the large role of the federal government. On the left, many teachers' unions have concerns about the testing components. And so, I think we need to make several positive changes. And I see several of those being made. I can see us making some improvements in the way we measure students with special education needs. I see some positive changes in the way we deal with children with limited English proficiency. I see the growth model being used at least as a supplement, if not more.

But the biggest remaining complaint that I hear about No Child Left Behind in Florida is the inconsistency between the state and the federal accountability systems. And I am very interested in hearing from you about how the states and the feds can better align their dual accountability systems to ensure that parents are given clear and consistent information about their children's schools.

Let me just give you an example. In Florida, we use one test called the FCAT both for the state's program called the A Plus program and for the federal No Child Left Behind program. Approximately 90 percent of the schools get a passing score under the state plan. And approximately 90 percent of the schools fail to meet AYP under the federal plan.

So a parent moves into a school district and says, "Is this a good school?" Well, it is failing under the federal program, and it is an A school under the state program. And I think we have got to bring those in line.

And so, I want to ask you. Let me start with Mr. McWalters.

Are you also concerned as we go through reauthorization about this all or nothing approach to measuring progress for AYP? And if so, do you think we should go with a more graduated approach in terms of bringing the states and the feds more in line?

Mr. MCWALTERS. More graduated. However, I want to go on record. I think the feds need to stay in this business. We wouldn't be having this conversation if states either had the capacity or the will five generations ago to get us to where we are now.

Mr. KELLER. Nobody is questioning that.

Mr. MCWALTERS. So having said that, now I am talking about the spirit of the law versus the way it is administered.

Mr. KELLER. But I have only got a limited amount of time. I just want that—do you think we should go to a more graduated approach instead of——

Mr. MCWALTERS. But it has to have a peer review structure that is transparent.

Mr. KELLER. Right.

Mr. MCWALTERS. Because when my proposal is being reviewed, it is being reviewed in a way that appreciates the context from which I am coming.

Mr. KELLER. Let me stop you there. I hate to, but I have just got a little amount of time.

Secretary Woodruff, do you believe that we should continue with this all or nothing sort of approach with AYP? Or would you prefer a more graduated approach?

Ms. WOODRUFF. Absolutely a more graduated approach.

Mr. KELLER. And do you have any ideas how states and the federal government can bring their dual accountability systems more in line?

Ms. WOODRUFF. Well, again, I think that, you know, in the reauthorization if you set some criteria around which—a framework within which we have to work and then allow us to bring forward our proposals that are measured then against this criteria, it makes sense.

In Delaware, for example, we use both AYP and a growth component for our school rating. And we use growth of all children at all levels in reading, mathematics, science, and social studies as a part of that because we continue to value all four content areas, not just reading and math.

Mr. KELLER. Well, I met with our local bureaucrats at our Florida Department of Education. I asked them how could we bring them in line. And they did the data analysis for me. And if you meet 90 percent of the AYP criteria and call that excellent, say, that equals almost identical to schools who get an a. If you meet 80 percent of the criteria, we will call that good. That meets almost identical the schools that get a B.

If you meet 70 percent and call it average, that meets almost identical the number of schools to get a C. But I am told when talking to folks on both sides of the aisle that if we did that sort of evaluative process on AYP that that would hurt some schools' feelings, that, you know, they are only average or good.

And so, let me ask you, Mr. Olson, do you like that sort of graduated approach. Or do you think we should stay with the all or nothing approach to AYP?

Mr. OLSON. I would prefer the graduated. I also think that it is important to maintain some of the richness that state systems have. And I wouldn't be a real strong fan of adding many additional measures inside the calculation of AYP. I think that the schools should have multiple measures. I think states have the position and obligation to put those in place.

And when you have a richer state system than you would want to fund and put in place from a federal system, I think you will have disparity from time to time. But I think the states have the flexibility also to create some means by which they appear more consistent.

Mr. KELLER. My time is expired unfortunately. I yield back.

Chairman MILLER. Mr. Payne?

Mr. PAYNE. Thank you very much.

You know, when this legislation first came—and like Mr. Courtney said, I was troubled because I knew that schools that had poor fiscal conditions, unqualified teachers, over-crowded classes, which are primarily in urban areas like mine in Newark, New Jersey and other urban places, I was somewhat opposed, disturbed by highstakes testing because I knew that they were going to show up at the bottom because of not having the opportunity to learn, which was a part of legislation in the past.

But the majority that was in control for the last 12 years took out opportunity to learn. So if you were failing, that is your problem. It wasn't that you were not provided with the opportunity to learn.

Secondly, I knew that there would be some problems with the suburban communities that might send large numbers of children to colleges. However, with No Child Left Behind it sort of disaggregated.

And therefore, you could see that there were children being left behind because this legislation showed that there were minority kids, English proficiency language and special needs kids who were being left behind by these school districts that sent the majority of their kids off to wherever they would go after high school, but there was very little acknowledgement for the others. So I was kind of conflicted with knowing the testing was going to show negatively, on the other hand, knowing that the testing would show that there were almost discrimination to other kids.

The whole question of states' rights—I mean, that is why we were so far behind. That is why we had to start with a national lunch program because states weren't taking care of people when World War II started. Title I, because they didn't deal with low-income school districts.

So the federal government said, well, put this in. And the states who still have some of those old trends about not wanting government to intervene is because of things like public accommodations, the old Jim Crow laws, the old voting rights. And they don't people to expose the discrimination that still exists.

Having said that, though, let us get back to the topic on-hand. Let me ask a quick question. First 3 years of No Child Left Behind, growth models were generally not considered to be consistent with certain statutory provisions of the law. However, as you all know, in 2005, the secretary of education reversed course and announced that a pilot project under which up to 10 states would be allowed to use growth models to make AYP determined for that school year of 2005, 2006.

Do you feel that the growth models overstate progress or appropriate credit improving schools? And you could also, if you have any comment or disagreement with my previous statement, you may certainly want to run in that all in about another 2 minutes.

Mr. OLSON. From what I have seen in the data, it does not seem to have any negative effects relative to the requirements of the law. There are relatively few schools that are making AYP with the growth model that weren't before. So that hasn't shifted much. I

think it is very important to know that it is important to measure growth just because it is the best indicator of effectiveness of systems.

I don't believe states are moving to measuring growth so fewer schools would be identified in that category. I haven't heard that in any of the conversations in any state. And I believe that they are functioning with a great deal of integrity. So I think it is all a positive move.

Ms. WOODRUFF. What I am going to be interested in seeing is that once we put this growth model in place and we have more definitive information that schools receive—I want to see then what the effect of that is and their ability to intervene and do more for the individual children and groups of children so that they are moving either out of school improvement or continue on a trajectory to continue to meet the target. So I think that will not be known until we see this over probably at least a 3-year term relative to the examination of the data and what happens. But it is not an attempt to duck the system at all.

Mr. DORAN. The growth models are entirely consistent with the idea of what it means to learn. When a kid is learning, we know that the student is growing and changing. And so, growth models, when properly developed, reflect that notion.

Dr. Dougherty and I serve on the secretary's peer review panel. And I think that panel worked very well in this last round. In fact, there were some growth models that statistically may have allowed for some schools to over-express growth. And they were met with some concern and comment from whether they were defensible or not.

And I think if these growth models are to be allowed, that this peer review process that scrutinized the statistical methods that were being used and whether they would do exactly what you are asking—would they over-credit schools—needs to be emphasized and needs to continue to be in place to guard against exactly the point that you are mentioning. I do think growth models should be applied because they are the right thing to do. But I also think they should be subject to statistical scrutiny and whether they fit reasonably within a policy context.

Mr. DOUGHERTY. And I will mention that there was a lot of conversation in the panel about over time validating the growth models to see how many of the kids who were predicted to be proficient are on track to proficient actually end up being proficient.

Mr. MCWALTERS. I obviously support growth models. But don't substitute the instrument of measurement for the causation of change. Your issues about concentrated student need—the growth model is just going to help us see it. It is not going to answer how you treat it.

Chairman MILLER. The gentleman's time is expired.

If I might follow on with a second round of questioning here, although I see we—excuse me. Mr. Ehlers? I am sorry.

Mr. EHLERS. Thank you.

Chairman MILLER. The gentleman is recognized.

Mr. EHLERS. As a token scientist here, I am used to being overlooked. But also as a token scientist, I have to ask a question about science education or my colleagues will think I have lost my ability.

At any rate, Dr. Dougherty, I noticed that you taught in elementary school, taught science. And you are aware, of course, that schools have to begin testing for science in 2007, 2008. But these tests under current law do not count toward AYP. I am proposing that they should. And I would appreciate your comment on that and whether you think that is an appropriate thing.

Mr. DOUGHERTY. I think they should. I think that—just going back to my experience, back in the day, a lot of times districts didn't have science curricula for elementary schools. In Texas, teachers, the science teachers actually requested that the tests count in the state accountability system because otherwise the school systems wouldn't pay enough attention to teaching science. So I think making science count is important.

Mr. EHLERS. I appreciate that. And I, in fact, have introduced a bill to add that to No Child Left Behind. I hope it is included in the reauthorization.

Let me go beyond that now. Some of you have made comments about the multitude of tests, the variability in the tests. My colleague who just left, Mr. Keller, raised the point that it was hard to keep track of who was doing well and who didn't because of the testing methods.

I have introduced a bill to provide voluntary educational standards, math and science standards. And schools would not be required to use them, but obviously we would encourage them to use them. And I have a reason for that. You might argue it would be better to have national standards in other areas, but certainly, in the science and math because it is sequential in nature. And because of the variability of textbooks, the schedules and coupled with the mobility of families and students in today's world, it is very possible for students to get messed up.

For example, if a student is attending a school that teaches fractions in the fall, percentages in the spring, and in January transfers to a school that teaches percentages in the fall, fractions in the spring, they get a double dose of fractions and never learn percentages. That is not an uncommon problem. I have seen it in a number of schools.

Do you think it makes sense that we have a system of voluntary standards? And particularly, this came about not because so much of the sequencing, but when I looked at test results and this recent comparison that came out, comparison between how students did on the NAEP test compared to how they did on the state's tests, my own state got a D-in terms of how well the students were performing on the NAEP test compared to how they performed on the state test. And Michigan is an outstanding state, has a good school system.

So there is something wrong if we don't have a better national standard so that we can compare apples to oranges related to AYP in different states. Any comments on that?

Mr. OLSON. I think maybe everyone on the panel will want to comment on it. Dr. Doran made a comment earlier that made reference to how states establish their benchmark, their requirement for proficiency. As far as I know, states have put proficiency statements in place that have no relative relationship to anything real in the world. NAEP is an example of that.

So if we do move to voluntary standards, which I would be in favor of personally, that we do it in such a way that we ask the question what is it in the real world that should create that anchor of expectation and make that common across the states. The NAEP standard probably is not that standard. And so, I would suggest some serious thought. And to the extent that common standards or voluntary standards spread across other academic areas, the same question would be raised.

Mr. EHLERS. That is a good idea, good comment.

Others? Yes?

Ms. WOODRUFF. I think that it absolutely is time for us to have voluntary national standards. And by that, I don't mean federal standards. I mean standards that we come together, we agree what the standards are. And we have to be thinking more clearly about serving the needs of our students, who are a much more mobile population today than they have ever been. So the conversation around national standards is timely, appropriate, and we ought to have it.

Mr. DOUGHERTY. I think such standards would be tremendously influential, which means if they are very good standards, very strong standards, they would be very positively influential. So it would be very, very important, particularly important, to get them right and have them to be strong. I suggest one of the anchors should be the aim that students be ready for college, and skill careers be a target for those standards.

Mr. MCWALTERS. I am from Rhode Island. So we have voluntary cooperative standards with two other states. I advocate it. I think it is got to be voluntarily. I am more interested in the measurements, the instrumentations of the standards and how we use measurement to actually get students to hit standards that are comprehensive. Don't confuse the standards with the need for multiple measures of them.

Mr. EHLERS. Thank you.

Dr. Doran, any comments?

Mr. DORAN. I do have comments. I mentioned this in my testimony and with relationship to the NAEP specifically. I do think—and I am a strong supporter—of voluntary national standards. I think the question is why do we have so much variability in the states' performance levels, and can we do a better job in bringing some coherence into our educational accountability system because of the reason that you mentioned, that we have a very mobile society. So for that reason, I am, in fact, very supportive of voluntary national standards.

I do want to dovetail on what Allan Olson mentioned a moment ago. And that is that if voluntary national standards are created, especially as we look toward the high school, those standards should begin the conversation of connecting those standards with skills required to be successful post-high school.

Mr. EHLERS. Thank you very much. I appreciate that.

Chairman MILLER. It probably would be too logical of a conclusion. But we will try it.

Let me just ask a question. I am sorry. We have a vote, and I don't want to hold you for that vote.

But, Mr. Dougherty, you indicated that there are 27 states that now have in place a data system that you think is acceptable so that they can move to a growth model. Is that a fair statement of your testimony?

Mr. DOUGHERTY. That is a fair statement. We didn't look at their assessment system, but we looked at their data system.

Chairman MILLER. So if the decision is made to go out and to embrace a growth model—and I assume we are all talking about a growth model toward proficiency, that this is a growth model to take you somewhere, that that is the kind of model. And there is obviously multiple growth models available, as I understand it, with integrity and with credibility for the results that we sort of have in this common conversation about what we want to achieve. So how do we start that transition? What do you do with it?

I notice my state is not on that list of a state that has a data systems acceptable. And they just got a report, just a huge report that they have waited 3 years for that essentially one of the components has told them that their data system is in a shambles. They really know very little about their customers at all, where they are, what they are doing or how they are coming and going.

What happens to them in this transition period? I mean, do we go through the process that we have been going through? You are on the secretary's peer review. States continue to make applications, and they are deemed adequate. And that is the process by which they get through.

And I don't know, Secretary Woodruff, if you have had your experience with that process.

But if you might, outline that, those who feel confident to do so.

Mr. DOUGHERTY. I would comment that is a very good process. It basically causes—it is voluntary. States step up to the plate. Everybody pretty much wants to have a growth model. And so, it is kind of like do you qualify.

Chairman MILLER. Yes, but a lot of people want a growth model because they think it is a silver bullet.

Mr. DOUGHERTY. Yes.

Chairman MILLER. You can hear it——

Mr. DOUGHERTY. It is not a silver bullet.

Chairman MILLER. You can hear it in their voice sometimes when they talk to you about it.

Mr. DOUGHERTY. They are going to be surprised how few additional schools qualify for AYP, as North Carolina, I think, has found, Delaware is likely to find. It is not a silver bullet. But from the point of view of improving the evaluation of the effectiveness of educational programs, providing guidance for school improvement, it is not necessarily a silver bullet. But it is definitely something that ought to be in the armory.

Chairman MILLER. Thank you.

Mr. Doran?

Mr. DORAN. I will follow that and say that I didn't serve on the first round of peer reviews, but I did serve on the most recent rounds. I like the process that is currently in place. So to get from A to B if the flexibility were awarded such that states could implement the growth model, I think those growth models need to be

submitted in the application process. This is very similar to the way states did this at the very beginning.

When NCLB was first established, they had to establish an accountability workbook, and they had to go through the process of how they were going to compute AYP and so forth. I would use that same process, that states would have to describe how they are going to implement the growth model, how they are going to use it within their accountability system. It should then be scrutinized, modified, if needed.

And I would also support the notion that it didn't turn out to be a silver bullet in either Tennessee or North Carolina. I think Tennessee had seven additional schools that made AYP as a result. And North Carolina, I believe, had none, if I remember my facts correctly.

Mr. DOUGHERTY. And in contract with the accountability workbook process, which is pretty much mandatory for 50 states, I wouldn't make having a growth model be a mandatory—you will get more enthusiastic participation if it is voluntary and probably more ingenuity of the ones who apply.

Chairman MILLER. Mr. Olson, let me ask you this. In your testimony you obviously lay out, you know, a substantial track record of looking at these systems and administering these longitudinal tests and the results. And you find this all compatible with your experience that states would be able to adapt to a system that would be able to allow them to mine this kind of information from these models that are—I guess I want to say—currently under consideration?

Mr. OLSON. Yes, I do. The thing I would come back to is that the states are allowed to assess even more accurately the wealth and the information and the value of the information will become increasingly useful and give us an opportunity to target and improve decision makings on many people inside the educational system in contrast to, you know, just the district level or just the state level.

Chairman MILLER. Let me ask you if you might, just quickly, what is the red flag we should be looking for in terms of when people describe to us the process they, their state, would like to go through to get to the other side. Is there a red flag that you have watched in the secretary's process or in experience of people who—I always worry that people embrace a concept but then their vision of the concept is a little skewed.

Mr. MCWALTERS. I think I can answer that from a state's perspective. If I have a growth model or not right now and I pass the review of the experts, my gap to 2014 is not going to get smaller with a growth model. So the issue of understanding how far we are as a nation from wrestling with proficiency at real levels without softening the bar—none of us want to soften the bar.

So when you have a growth model or not, the gap is real. And the intervention capacity question is still the part that is missing for me. I don't want to hide how far I have to go. I want to change my capacity to get there.

Chairman MILLER. Anyone else?

Secretary?

Ms. WOODRUFF. As far as the whole growth model issue is concerned, I think that it is very important that the whole process is clear, understandable, and transparent.

Chairman MILLER. That is the congressional process.

Ms. WOODRUFF. So that there is absolutely no question about what the criteria are, how they are going to be judged, and that the conversation is iterative. And as far as I am concerned, if there are 27 states——

Chairman MILLER. You are talking about the approval process for that growth model.

Ms. WOODRUFF. I am talking about what is it you have to do and what are the steps that must be taken and then how are you going to be judged. I don't want to know what the test is after I have taken the course. I would like to know ahead of time what I am going to be judged on. And I think that has been the concern that a number of us had.

If there are 27 states ready now, let them go. And then we will help the other states understand what the mechanisms are and the hurdles are to get there. I think we are in a state of today where nationally we really help each other and step up to do that on a regular basis.

Chairman MILLER. Thank you. That may be a good place to interrupt this conversation. I hope that we will be able to continue it as the committee gets deeper into the reauthorization process.

Thank you so much for your time and your expertise and your experiences. I think this was very, very helpful to the members of the committee.

The hearing record will stay open for 14 days. If there are others who want to make submissions, we would certainly take them under consideration.

[The prepared statement of Mr. Altmire follows:]

Prepared Statement of Hon. Jason Altmire, a Representative in Congress From the State of Pennsylvania

Thank you, Mr. Chairman, for holding this hearing to examine how we can improve No Child Left Behind's measures of progress.

I would like to extend a warm welcome to today's witnesses. I appreciate all of you for taking the time to be here and look forward to hearing from you.

Measuring whether or not students are making Adequate Yearly Progress is fundamental to how NCLB works. We must have indicators that accurately measure student knowledge and track their academic achievement to determine which schools are truly in need of intervention and to determine exactly what interventions are needed.

I am particularly interested in hearing our witnesses' comments on growth models. Pennsylvania's proposal to institute a growth-based accountability model has just begun the peer review process. Assessing student achievement in this way may have the potential to improve how we measure Adequate Yearly Progress because it allows for the tracking of individual students' academic gain on a yearly basis. However, I am aware that there are different types of growth models and would be interested in hearing about the best practices in this area.

Thank you again, Mr. Chairman. I yield back the balance of my time.

[Additional materials submitted by Chairman Miller follow:]

[The prepared statement of Prof. Hammond follows:]

Prepared Statement of Linda Darling-Hammond, Charles E. Ducommun Professor, Stanford University School of Education

I thank Chairman Miller and the members of the Committee for the opportunity to offer testimony on the re-authorization of ESEA, in particular the ways in which we measure and encourage school progress and improvement. My perspective on these issues is informed by my research, my work with states and national organizations on standards development, and my work with local schools. I have studied the implementation of No Child Left Behind,[1] as well as testing and accountability systems within the United States and abroad.[2] I have also served as past Chair of the New York State Council on Curriculum and Assessment and of the Chief State School Officers' INTASC Standards Development Committee. I work closely with a number of school districts and local schools on education improvement efforts, including several new urban high schools that I have helped to launch. Thus, I have encountered the issues of school improvement from both a system-wide and local school vantage point.

I am hopeful that this re-authorization can build on the strengths and opportunities offered by No Child Left Behind, while addressing needs that have emerged during the first years of the law's implementation. Among the strengths of the law is its focus on improving the academic achievement of all students, which triggers attention to school performance and to the needs of students who have been underserved, and its insistence that all students are entitled to qualified teachers, which has stimulated recruitment efforts in states where many disadvantaged students previously lacked this key resource for learning.

The law has succeeded in getting states, districts, and local schools to pay attention to achievement. The next important step is to ensure that the range of things schools and states pay attention to actually helps them improve both the quality of education they offer to every student and the quality of the overall schooling enterprise. In order to accomplish this, I would ask you to actively encourage states to:

- Develop accountability systems that use multiple measures of learning and other important aspects of school performance in evaluating school progress;
- Differentiate school improvement strategies for schools based on a comprehensive analysis of their instructional quality and conditions for learning.

Why Use Multiple Measures?

There are at least three reasons to gauge student and school progress based on multiple measures of learning and school performance:

- To direct schools' attention and effort to the range of measures that are associated with high-quality education and improvement;
- To avoid dysfunctional consequences that can encourage schools, districts, or states to emphasize one important outcome at the expense of another; for example, focusing on a narrow set of skills at the expense of others that are equally critical, or boosting test scores by excluding students from school; and
- To capture an adequate and accurate picture of student learning and attainment that both measures and promotes the kinds of outcomes we need from schools.

Directing Attention to Measures Associated with School Quality

One of the central concepts of NCLB's approach is that schools and systems will organize their efforts around the measures for which they are held accountable. Because attending to any one measure can be both partial and problematic, the concept of multiple measures is routinely used by policymakers to make critical decisions about such matters as employment and economic forecasting (for example, the Dow Jones Index or the GNP) and admission to college, where grades, essays, activities, and accomplishments are considered along with test scores.

Successful businesses use a "dashboard" set of indicators to evaluate their health and progress, aware that no single indicator is sufficient to understand or guide their operations. This approach is designed to focus attention on those aspects of the business that describe elements of the business's current health and future prospects, and to provide information that employees can act on in areas that make a difference for improvement. So, for example, a balanced scorecard is likely to include among its financial indicators not only a statement of profits, but also cash flow, dividends, costs and accounts receivable, assets, inventory, and so on. Business leaders understand that efforts to maximize profits alone could lead to behaviors that undermine the long-term health of the enterprise.

Similarly, a single measure approach in education creates some unintended negative consequences and fails to focus schools on doing those things that can improve their long-term health and the education of their students. Although No Child Left Behind calls for multiple measures of student performance, the implementation of the law has not promoted the use of such measures for evaluating school progress.

As I describe in the next section, the focus on single, often narrow, test scores in many states has created unintended negative consequences for the nature of teaching and learning, for access to education for the most vulnerable students, and for the appropriate identification of schools that are in need of improvement.

A multiple measures approach that incorporates the right "dashboard" of indicators would support a shift toward "holding states and localities accountable for making the systemic changes that improve student achievement" as has been urged by the Forum on Education and Accountability. This group of 116 education and civil rights organizations—which include the National Urban League, NAACP, League of United Latin American Citizens, Aspira, Children's Defense Fund, National Alliance of Black School Educators, and Council for Exceptional Children, as well as the National School Boards Association, National Education Association, and American Association of School Administrators—has offered a set of proposals for NCLB that would focus schools, districts, and states on developing better teaching, a stronger curriculum, and supports for school improvement.

Avoiding Dysfunctional Consequences

Another reason to use a multiple measures approach is to avoid the negative consequences that occur when one measure is used to drive organizational behavior.

The current accountability provisions of the Act, which are focused almost exclusively on school average scores on annual tests, actually create large incentives for schools to keep students out and to hold back or push out students who are not doing well. A number of studies have found that systems that reward or sanction schools based on average student scores create incentives for pushing low-scorers into special education so that their scores won't count in school reports,[3] retaining students in grade so that their grade-level scores will look better,[4] excluding low-scoring students from admissions,[5] and encouraging such students to leave schools or drop out.[6]

Studies in New York,[7] Texas,[8] and Massachusetts,[9] among others, have showed how schools have raised their test scores while "losing" large numbers of low-scoring students. For example, a recent study in a large Texas city found that student dropouts and push outs accounted for most of the gains in high school student test scores, especially for minority students. The introduction of a high-stakes test linked to school ratings in the 10th grade led to sharp increases in 9th grade student retention and student dropout and disappearance. Of the large share of students held back in the 9th grade, most of them African American and Latino, only 12% ever took the 10th grade test that drove school rewards. Schools that retained more students at grade 9 and lost more through dropouts and disappearances boosted their accountability ratings the most. Overall, fewer than half of all students who started 9th grade graduated within 5 years, even as test scores soared.[10]

Paradoxically, NCLB's requirement for disaggregating data and tracking progress for each subgroup of students increases the incentives for eliminating those at the bottom of each subgroup, especially where schools have little capacity to improve the quality of services such students receive. Table 1 shows how this can happen. At "King Middle School," average scores increased from the 70th to the 72nd percentile between the 2002 and 2003 school year, and the proportion of students in attendance who met the proficiency standard (a score of 65) increased from 66% to 80%—the kind of performance that a test-based accountability system would reward. Looking at subgroup performance, the proportion of Latino students meeting the standard increased from 33% to 50%, a steep increase.

However, not a single student at King improved his or her score between 2002 and 2003. In fact, the scores of every single student in the school went down over the course of the year. How could these steep improvements in the school's average scores and proficiency rates have occurred? A close look at Table 1 shows that the major change between the two years was that the lowest-scoring student, Raul, disappeared. As has occurred in many states with high stakes-testing programs, students who do poorly on the tests—special needs students, new English language learners, those with poor attendance, health, or family problems—are increasingly likely to be excluded by being counseled out, transferred, expelled, or by dropping out.

TABLE 1.—KING MIDDLE SCHOOL: REWARDS OR SANCTIONS?

[The Relationship between Test Score Trends and Student Populations]

	2002–03	2003–04
Laura	100	90
James	90	80
Felipe	80	70

TABLE 1.—KING MIDDLE SCHOOL: REWARDS OR SANCTIONS?—Continued

[The Relationship between Test Score Trends and Student Populations]

	2002–03	2003–04
Kisha	70	65
Jose	60	55
Raul	20	
	Ave. Score = 70% meeting standard = 66%	Ave. Score = 72% meeting standard = 80%

This kind of result is not limited to education. When one state decided to rank cardiac surgeons based on their mortality rates, a follow up investigation found that surgeons' ratings went up as they stopped taking on high-risk clients. These patients were referred out of state if they were wealthy, or were not served, if they were poor.

The three national professional organizations of measurement experts have called attention to such problems in their joint Standards for Educational and Psychological Testing, which note that:

Beyond any intended policy goals, it is important to consider potential unintended effects that may result from large-scale testing programs. Concerns have been raised, for instance, about narrowing the curriculum to focus only on the objectives tested, restricting the range of instructional approaches to correspond to the testing format, increasing the number of dropouts among students who do not pass the test, and encouraging other instructional or administrative practices that may raise test scores without affecting the quality of education. It is important for those who mandate tests to consider and monitor their consequences and to identify and minimize the potential of negative consequences.[11]

Professional testing standards emphasize that no test is sufficiently reliable and valid to be the sole source of important decisions about student placements, promotions, or graduation, but that such decisions should be made on the basis of several different kinds of evidence about student learning and performance in the classroom. For example, Standard 13.7 states:

In educational settings, a decision or characterization that will have major impact on a student should not be made on the basis of a single test score. Other relevant information should be taken into account if it will enhance the overall validity of the decision.[12]

The Psychological Standards for Testing describe several kinds of information that should be considered in making judgments about what a student knows and can do, including alternative assessments that provide other information about performance and evidence from samples of school work and other aspects of the school record, such as grades and classroom observations. These are particularly important for students for whom traditional assessments are not generally valid, such as English language learners and special education students. Similarly, when evaluating schools, it is important to include measures of student progress through school, coursework and grades, and graduation, as part of the record about school accomplishments.

Evaluating Learning Well

Indicators beyond a single test score are important not only for reasons of validity and fairness in making decisions, but also to assess important skills that most standardized tests do not measure. Current accountability reforms are based on the idea that standards can serve as a catalyst for states to be explicit about learning goals, and the act of measuring progress toward meeting these standards is an important force toward developing high levels of achievement for all students. However, an on-demand test taken in a limited period of time on a single day cannot measure all that is important for students to know and be able to do. A credible accountability system must rest on assessments that are balanced and comprehensive with respect to state standards. Multiple-choice and short-answer tests that are currently used to measure standards in many states do not adequately measure the complex thinking, communication, and problem solving skills that are represented in national and state content standards.

Research on high-stakes accountability systems shows that, "what is tested is what is taught," and those standards that are not represented on the high stakes assessment tend to be given short shrift in the curriculum.[13] Students are less likely to engage in extended research, writing, complex problem-solving, and experimentation when the accountability system emphasizes short-answer responses to formulaic problems. These higher order thinking skills are those very skills that often are cited as essential to maintaining America's competitive edge and necessary

for succeeding on the job, in college, and in life. As described by Achieve, a national organization of governors, business leaders, and education leaders, the problem with measures of traditional on-demand tests is that they cannot measure many of the skills that matter most for success in the worlds of work and higher education:

States * * * will need to move beyond large-scale assessments because, as critical as they are, they cannot measure everything that matters in a young person's education. The ability to make effective oral arguments and conduct significant research projects are considered essential skills by both employers and postsecondary educators, but these skills are very difficult to assess on a paper-and pencil test.[14]

One of the reasons that U.S. students fall further and further behind their international counterparts as they go through school is because of differences in curriculum and assessment systems. International studies have found that the U.S. curriculum focuses more on superficial coverage of too many topics, without the kinds of in-depth study, research, and writing needed to secure deep understanding. To focus on understanding, the assessment systems used in most high-achieving countries around the world emphasize essay questions, research projects, scientific experiments, oral exhibitions and performances that encourage students to master complex skills as they apply them in practice, rather than multiple-choice tests.

As indicators of the growing distance between what our education system emphasizes and what leading countries are accomplishing educationally, the U.S. currently ranks 28th of 40 countries in the world in math achievement—right above Latvia—and 19th of 40 in reading achievement on the international PISA tests that measure higher-order thinking skills. And while the top-scoring nations—including previously low-achievers like Finland and South Korea—now graduate more than 95% of their students from high school, the U.S. is graduating about 75%, a figure that has been stagnant for a quarter century and, according to a recent ETS study, is now declining. The U.S. has also dropped from 1st in the world in higher education participation to 13th, as other countries invest more resources in their children's futures.

Most high-achieving nations' examination systems include multiple samples of student learning at the local level as well as the state or national level. Students' scores are a composite of their performance on examinations they take in different content areas—featuring primarily open-ended items that require written responses and problem solutions—plus their work on a set of classroom tasks scored by their teachers according to a common set of standards. These tasks require them to conduct apply knowledge to a range of tasks that represent what they need to be able to do in different fields: find and analyze information, solve multi-step real-world problems in mathematics, develop computer models, demonstrate practical applications of science methods, design and conduct investigations and evaluate their results, and present and defend their ideas in a variety of ways. Teaching to these assessments prepares students for the real expectations of college and of highly skilled work.

These assessments are not used to rank or punish schools, or to deny promotion or diplomas to students. In fact, several countries have explicit proscriptions against such practices. They are used to evaluate curriculum and guide investments in professional learning—in short, to help schools improve. By asking students to show what they know through real-world applications of knowledge, these nations' assessment systems encourage serious intellectual activities on a regular basis. The systems not only measure important learning, they help teachers learn how to design curriculum and instruction to accomplish this learning.

It is worth noting that a number of states in the U.S. have developed similar systems that combine evidence from state and local standards-based assessments to ensure that multiple indicators of learning are used to make decisions about individual students and, sometimes, schools. These include Connecticut, Kentucky, Maine, Nebraska, New Hampshire, Oregon, Rhode Island, Pennsylvania, Vermont, and Wyoming, among others. However, many of these elements of state systems are not currently allowed to be used to gauge school progress under NCLB.

Encouraging these kinds of practices could help improve learning and guide schools toward more productive instruction. Studies have found that performance assessments that are administered and scored locally help teachers better understand students' strengths, needs, and approaches to learning, as well as how to meet state standards.[15] Teachers who have been involved in developing and scoring performance assessments with other colleagues have reported that the experience was extremely valuable in informing their practice. They report changes in both the curriculum and their instruction as a result of thinking through with colleagues what good student performance looks like and how to better support student learning on specific kinds of tasks.

These goals are not well served by external testing programs that send secret, secured tests into the school and whisk them out again for machine scoring that produces numerical quotients many months later. Local performance assessments provide teachers with much more useful classroom information as they engage teachers in evaluating how and what students know and can do in authentic situations. These kinds of assessment strategies create the possibility that teachers will not only teach more challenging performance skills but that they will also be able to use the resulting information about student learning to modify their teaching to meet the needs of individual students. Schools and districts can use these kinds of assessments to develop shared expectations and create an engine for school improvement around student work.

Research on the strong gains in achievement shown in Connecticut, Kentucky, and Vermont in the 1990s attributed these gains in substantial part to these states' performance-based assessment systems, which include such local components, and related investments in teaching quality.[16] Other studies in states like California, Maine, Maryland, and Washington,[17] found that teachers assigned more ambitious writing and mathematical problem solving, and student performance improved, when assessments included extended writing and mathematics portfolios and performance tasks. Encouraging these kinds of measures of student performance is critical to getting the kind of learning we need in schools.

Not incidentally, more authentic measures of learning that go beyond on-demand standardized tests to look directly at performance are especially needed to gain accurate measures of achievement for English language learners and special needs students for whom traditional tests are least likely to provide valid measures of understanding.[18]

What Indicators Might be Used to Gauge School Progress?

A key issue is what measures should be used to determine Adequate Yearly Progress (AYP) or the alternative tools that are used for addressing NCLB's primary goals, e.g. assuring high expectations for all students, and helping schools address the needs of all students. Current AYP measures are too narrow in several respects: They are based exclusively on tests which are often not sufficient measures of our educational goals; they ignore other equally important student outcomes, including staying in school and engaging in rigorous coursework; they ignore the growth made by students who are moving toward but not yet at a proficiency benchmark, as well as the gains made by students who have already passed the proficiency benchmark; and they do not provide information or motivation to help schools, districts, and states improve critical learning conditions.

This analysis suggests that school progress should be evaluated on multiple measures of student learning—including local and state performance assessments that provide evidence about what students can actually do with their knowledge—and on indicators of other student outcomes, including such factors as student progress and continuation through school, graduation, and success in rigorous courses. The importance of these indicators is to encourage schools to keep students in school and provide them with high-quality learning opportunities—elements that will improve educational opportunities and attainment, not just average test scores.

To these two categories of indicators, I would add indicators of learning conditions that point attention to both learning opportunities available to students (e.g. rigorous courses, well-qualified teachers) and to how well the school operates. In the business world, these kinds of measures are called leading indicators, which represent those things that employees can control and improve upon. These typically include evidence of customer satisfaction, such as survey data, complaints and repeat orders; as well as of employee satisfaction and productivity, such as employee turnover, project delays, evidence of quality and efficiency in getting work done; reports of work conditions and supports, and evidence of product quality.

Educational versions of these kinds of indicators are available in many state accountability systems. For example, State Superintendent Peter McWalters noted in his testimony to this committee that Rhode Island uses several means to measure school learning conditions. Among them is an annual survey to all students, teachers, and parents that provides data on "Learning Support Indicators" measuring school climate, instructional practices, and parental involvement. In addition, Rhode Island, like many other states, conducts visits to review every school in the state every five years, not unlike the Inspectorate system that is used in many other countries. These kinds of reviews can examine teaching practices, the availability and equitable allocation of school resources, and the quality of the curriculum, as it is enacted.

Ideally, evaluation of school progress would be based on a combination of these three kinds of measures and would emphasize gains and improvement over time,

both for the individual students in the school and for the school as a whole. Along with data about student characteristics, an indicator system could include:

- Measures of student learning: both state tests and local assessments, including performance measures that assess higher-order thinking skills and understanding, including student work samples, projects, exhibitions, or portfolios.
- Measures of additional student outcomes: data about attendance, student grade-to-grade progress (promotion / retention rates) and continuation through school (ongoing enrollment), graduation, and course success (e.g. students enrolled in, passing, and completing rigorous courses of study).
- Measures of learning conditions, data about school capacity, such as teacher and other staff quality, availability of learning materials, school climate (gauged by students', parents', and teachers' responses to surveys), instructional practices, teacher development, and parental engagement.

These elements should be considered in the context of student data, including information about student mobility, health, and welfare (poverty, homelessness, foster care, health care), as well as language background, race / ethnicity, and special learning needs—not a basis for accepting differential effort or outcomes, but as a basis for providing information needed to interpret and improve schools' operations and outcomes.

How Might Indicators be Used to Determine School Progress and Improvement Strategies?

The rationale for these multiple indicators is to build a more powerful engine for educational improvement by understanding what is really going on with students and focusing on the elements of the system that need to change if learning is to improve. High-performing systems need a regular flow of useful information to evaluate and modify what they are doing to produce stronger results. State and local officials need a range of data to understand what is happening in schools and what they should do to improve outcomes. Many problems in local schools are constructed or constrained by district and state decisions that need to be highlighted along with school-level concerns. Similarly, at the school level, teachers and leaders need information about how they are doing and how their students are doing, based in part on high-quality local assessments that provide rich, timely insights about student performance.

Some states and districts have successfully put some of these indicators in place. The federal government could play a leadership role by not only encouraging multiple measures for assessing school progress and conditions for learning but by providing supports for states to build comprehensive databases to track these indicators over time, and to support valid, comprehensive information systems at all levels.[19]

If we think comprehensively about the approach to evaluation that would encourage fundamental improvements in schools, several goals emerge. First, determinations of school progress should reflect an analysis of schools' performance and progress along several key dimensions. Student learning should be evaluated using multiple measures that provide comprehensive and valid information for all subpopulations. Targets should be based on sensible goals for student learning, examining growth from where students start, setting growth targets in relation to that starting point, and pegging "proficiency" at a level that represents a challenging but realistic standard, perhaps at the median of current state proficiency standards. Targets should also ensure appropriate assessment for special education students and English language learners and credit for the gains these students make over time. And analysis of learning conditions including the availability of materials, facilities, curriculum opportunities, teaching, and leadership should accompany assessments of student learning.

A number of states already have developed comprehensive indicator systems that can be sources of such data, and the federal government should encourage states to propose different means for how to aggregate and combine these data. In addition, many states' existing assessment systems already provide different ways to score and combine state reference tests with local testing systems, locally administered performance tasks (which are often scored using state standards), and portfolios.[20]

For evaluating annual progress, one likely approach would be to use an index of indicators, such as California's Academic Performance Index, which can include a weighted combination of data about state and local tests and assessments as well as other student outcome indicators like attendance, graduation, promotion rates, participation and pass rates or grades for academic courses. Assessment data from multiple sources and evidence of student progression through / graduation from school would be required components. Key conditions of learning, such as teacher qualifications, might also be required. Other specific indicators might be left to

states, along with the decision of how much weight to give each component, perhaps within certain parameters (for example, that at least 50 percent of a weighted index would reflect the results of assessment data).

Within this index, disaggregated data by race/ethnicity and income could be monitored on the index score, or on components of the overall index, so that they system pays ongoing attention to progress for groups of students. Wherever possible these measures should look at progress of a constant cohort of students from year to year, so that actual gains are observed, rather than changes in averages due to changes in the composition of the student population. Furthermore, gains for English language learners and special education students should be evaluated on a growth model that ensures appropriate testing based on professional standards and measures individual student growth in relation to student starting points.

Non-academic measures such as improved learning climate (as measured by standard surveys, for example, to allow trend analysis over time), instructional capacity (indicators regarding the quality of curriculum, teaching, and leadership), resources, and other contributors to learning could be included in a separate index on Learning Conditions, on which progress is also evaluated annually as part of both school, district, and state assessment.

Once school progress indicators are available, a judgment must be made about whether a school has made adequate progress on the index or set of indicators. If the law is to focus on supporting improvement it will be important to look at continuous progress for all students in a school rather than the "status model" that has been used in the past. A progress model would recognize the reasonable success of schools that deserve it. Rather than identifying a school as requiring intervention when a single target is missed (for example, if 94% of economically disadvantaged students take the mathematics test one year instead of 95%), a progress model would gauge whether the overall index score increases, with the proviso that the progress of key subgroups continues to be examined, with lack of progress a flag for intervention.

The additional use of the indicators schools and districts have assembled would be in the determination of what kind of action is needed if a school does not make sufficient progress in a year. To use resources wisely, the law should establish a graduated system of classification for schools and districts based on their rate of progress, ranging from state review to corrective actions to eventual reconstitution if such efforts fail over a period of time. States should identify schools and districts as requiring intervention based both on information about the overall extent of progress from the prior year(s) and on information about specific measures in the system of indicators—for example, how many progress indicators have lagged for how long. This additional scrutiny would involve a school review by an expert team—much like the inspectorate systems in other countries—that conducts an inspection of the school or LEA and analyzes a range of data, including evidence of individual and collective student growth or progress on multiple measures; analysis of student needs, mobility, and population changes; and evaluation of school practices and conditions. Based on the findings of this review, a determination would be made about the nature of the problem and the type of school improvement plan needed. The law should include the explicit expectation that state and district investments in ensuring adequate conditions for learning must be part of this plan.

The overarching goal of the ESEA should be to improve the quality of education students receive, especially those traditionally least well served by the current system. To accomplish this, the measures used to gauge school progress must motivate continuous improvement and attend to the range of school outcomes and conditions that are needed to ensure that all students are educated to higher levels.

ENDNOTES

[1] See, e.g. L. Darling-Hammond, No Child Left Behind and High School Reform, Harvard Education Review, 76, 4 (Winter 2006), pp. 642-667. *http://www.edreview.org/harvard06/2006/wi06/w06darli.htm*

L. Darling-Hammond, From 'Separate but Equal' to 'No Child Left Behind': The Collision of New Standards and Old Inequalities. In Deborah Meier and George Wood (eds.), Many Children Left Behind, pp. 3-32. NY: Beacon Press, 2004.

[2] Linda Darling-Hammond, Elle Rustique-Forrester, & Raymond Pecheone (2005). Multiple measures approaches to high school graduation: A review of state student assessment policies. Stanford, CA: Stanford University, School Redesign Network.

[3] Allington, R. L. & McGill-Franzen, A. (1992). Unintended effects of educational reform in New York, Educational Policy, 6 (4): 397-414; Figlio, D.N. & Getzler, L.S. (2002, April). Accountability, ability, and disability: Gaming the system? National Bureau of Economic Research.

[4] W. Haney (2000). The myth of the Texas miracle in education. Education Policy Analysis Archives, 8 (41): Retrieved Jul. 23, 07 from: *http://epaa.asu.edu/epaa/v8n41/*

[5] Smith, F., et al. (1986). High school admission and the improvement of schooling. NY: New York City Board of Education; Darling-Hammond, L. (1991). The Implications of Testing Policy for Quality and Equality, Phi Delta Kappan, November 1991: 220-225; Heilig, J. V. (2005), An analysis of accountability system outcomes. Stanford University.

[6] For recent studies examining the increases in dropout rates associated with high-stakes testing systems, see Advocates for Children (2002). Pushing out at-risk students: An analysis of high school discharge figures—a joint report by AFC and the Public Advocate. *http://www.advocatesforchildren.org/pubs/pushout-11-20-02.html;* W. Haney (2002). Lake Wobegone guaranteed: Misuse of test scores in Massachusetts, Part 1. Education Policy Analysis Archives, 10(24). *http://epaa.asu.edu/epaa/v10n24/;* J. Heubert & R. Hauser (eds.) (1999). High stakes: Testing for tracking, promotion, and graduation. A report of the National Research Council. Washington, D.C.: National Academy Press; B.A. Jacob (2001). Getting tough? The impact of high school graduation exams. Education and Evaluation and Policy Analysis 23 (2): 99-122; D. Lilliard, & P. DeCicca (2001). Higher standards, more dropouts? Evidence within and across time. Economics of Education Review, 20(5): 459-73;G. Orfield, D. Losen, J. Wald, & C.B. Swanson (2004). Losing our future: How minority youth are being left behind by the graduation rate crisis. Retrieved July 23, 2007 from: *http://www.urban.org/url.cfm?ID=410936;* M. Roderick, A.S. Bryk, B.A. Jacob, J.Q. Easton, & E. Allensworth (1999). Ending social promotion: Results from the first two years. Chicago: Consortium on Chicago School Research; R. Rumberger & K. Larson (1998). Student mobility and the increased risk of high school dropout. American Journal of Education, 107: 1-35; E. Rustique-Forrester (in press). Accountability and the pressures to exclude: A cautionary tale from England. Education Policy Analysis Archives; A. Wheelock (2003). School awards programs and accountability in Massachusetts.

[7] Advocates for Children (2002), Pushing out at-risk students; Heilig (2005), An analysis of accountability system outcomes; Wheelock (2003), School awards programs and accountability.

[8] Heilig, 2005.

[9] Wheelock, 2003

[10] Heilig, 2005.

[11] American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, Standards for Educational and Psychological Testing, Washington DC: American Educational Research Association, 1999, p.142.

[12] AERA, APA, NCME, Standards for Educational and Psychological Testing., p.146.

[13] See for example, Haney (2000). The myth of the Texas miracle; J.L. Herman & S. Golan (1993). Effects of standardized testing on teaching and schools. Educational Measurement: Issues and Practice, 12(4): 20-25, 41-42; B.D. Jones & R. J. Egley (2004). Voices from the frontlines: Teachers' perceptions of high-stakes testing. Education Policy Analysis Archives, 12 (39). Retrieved August 10, 2004 from *http://epaa.asu.edu/epaa/v12n39/;* M.G. Jones, B.D. Jones, B. Hardin, L. Chapman, & T. Yarbrough (1999). The impact of high-stakes testing on teachers and students in North Carolina. Phi Delta Kappan, 81(3): 199-203; Klein, S.P., Hamilton, L.S., McCaffrey, D.F., & Stetcher, B.M. (2000). What do test scores in Texas tell us? Santa Monica: The RAND Corporation; D. Koretz & S. I. Barron (1998). The validity of gains on the Kentucky Instructional Results Information System (KIRIS). Santa Monica, CA: RAND, MR-1014-EDU; D. Koretz, R.L. Linn, S.B. Dunbar, & L.A. Shepard (1991, April). The effects of high-stakes testing: Preliminary evidence about generalization across tests, in R. L. Linn (chair), The Effects of high stakes testing. Symposium presented at the annual meeting of the American Educational Research Association and the National Council on Measurement in Education, Chicago; R.L. Linn (2000). Assessments and accountability. Educational Researcher, 29 (2), 4-16; R.L. Linn, M.E. Graue, & N.M. Sanders (1990). Comparing state and district test results to national norms: The validity of claims that "everyone is above average." Educational Measurement: Issues and Practice, 9, 5-14; W. J. Popham (1999). Why Standardized Test Scores Don't Measure Educational Quality. Educational Leadership, 56(6): 8-15; M.L. Smith (2001). Put to the test: The effects of external testing on teachers. Educational Researcher, 20(5): 8-11.

[14] Achieve, Do graduation tests measure up? A closer look at state high school exit exams. Executive summary. Washington, DC: Achieve, Inc.

[15] L. Darling-Hammond & J. Ancess (1994). Authentic assessment and school development. NY: National Center for Restructuring Education, Schools, and Teaching, Teachers College, Columbia University; B. Falk & S. Ort (1998, September). Sitting down to score: Teacher learning through assessment. Phi Delta Kappan, 80(1): 59-64. G.L. Goldberg & B.S. Rosewell (2000). From perception to practice: The impact of teachers' scoring experience on the performance based instruction and classroom practice. Educational Assessment, 6: 257-290; R. Murnane & F. Levy (1996). Teaching the new basic skills. NY: The Free Press.

[16] J.B. Baron (1999). Exploring high and improving reading achievement in Connecticut. Washington: National Educational Goals Panel. Murnane & Levy (1996); B.M. Stecher, S. Barron, T. Kaganoff, & J. Goodwin (1998). The effects of standards-based assessment on classroom practices: Results of the 1996-97 RAND survey of Kentucky teachers of mathematics and writing. CSE Technical Report. Los Angeles: UCLA National Center for Research on Evaluation, Standards, and Student Testing; S. Wilson, L. Darling-Hammond, & B. Berry (2001). A case of successful teaching policy: Connecticut's long-term efforts to improve teaching and learning. Seattle: Center for the Study of Teaching and Policy, University of Washington.

[17] C. Chapman (1991, June). What have we learned from writing assessment that can be applied to performance assessment?. Presentation at ECS/CDE Alternative Assessment Conference, Breckenbridge, CO; J.L.Herman, D.C. Klein, T.M. Heath, S.T. Wakai (1995). A first look: Are claims for alternative assessment holding up? CSE Technical Report. Los Angeles: UCLA National Center for Research on Evaluation, Standards, and Student Testing; D. Koretz, K., J. Mitchell, S.I. Barron, & S. Keith (1996). Final Report: Perceived effects of the Maryland school performance assessment program CSE Technical Report. Los Angeles: UCLA National Center for Research on Evaluation, Standards, and Student Testing; W.A. Firestone, D. Mayrowetz, & J. Fairman (1998, Summer). Performance-based assessment and instructional

change: The effects of testing in Maine and Maryland. Educational Evaluation and Policy Analysis, 20: 95-113; S. Lane, C.A. Stone, C.S. Parke, M.A. Hansen, & T.L. Cerrillo (2000, April). Consequential evidence for MSPAP from the teacher, principal and student perspective. Paper presented at the annual meeting of the National Council on Measurement in Education, New Orleans, LA; B. Stecher, S. Baron, T. Chun, T., & K. Ross (2000) The effects of the Washington state education reform on schools and classroom. CSE Technical Report. Los Angeles: UCLA National Center for Research on Evaluation, Standards, and Student Testing.

[18] Darling-Hammond, Rustique-Forrester, and Pecheone, Multiple Measures.

[19] M. Smith paper (2007). Standards-based education reform: What we've learned, where we need to go. Consortium for Policy Research in Education.

[20] At least 27 states consider student academic records, coursework, portfolios of student work, and performance assessments, like research papers, scientific experiments, essays, and senior projects in making the graduation decision. Darling-Hammond, Rustique-Forrester, and Pecheone, Multiple Measures.

[National School Boards Association (NSBA) letter follows:]

March 20, 2007.

Hon. GEORGE MILLER, Chair,
Committee on Education and Labor, U.S. House of Representatives, Washington, DC.

Re: Hearing of the House Education and Labor Committee on Adequate Yearly Progress, March 21, 2007; National School Boards Association Statement for the Record.

DEAR CHAIRMAN MILLER: The National School Boards Association (NSBA), representing over 95,000 local school board members across the nation, commends you for your strong support to reauthorize the Elementary and Secondary Education Act (ESEA)/No Child Left Behind (NCLB) Act during the 110th Congress, and for establishing an aggressive schedule for congressional hearings over the coming weeks. NSBA looks forward to participating in future hearings and very much appreciates the opportunity to submit written testimony for the record.

Local school boards across the nation continue to support the goals of NCLB—including increased accountability for student performance. However, of utmost concern to local school boards is the belief that the current accountability framework does not accurately or fairly assess student, school, or school district performance.

Although the sponsors of the No Child Left Behind Act intended to establish a responsive accountability system for the nation's public schools, what has evolved in the name of accountability is a measurement framework that bases its assessment of school quality on a student's performance on a single assessment; and mandates a series of overbroad sanctions not always targeted to the students needing the services.

Five years after enactment of the federal law, local school districts continue to struggle to comply with the language of the law at a time when the unintended consequences of this complex law are imposing far more dysfunctional and illogical implementation problems than had been anticipated by the sponsors of the legislation. NSBA believes that the NCLB law can be amended to improve the accountability system in a way that restores public confidence in the law and results in significant improvement in the academic achievement of all students.

In January 2005, NSBA officially unveiled its bill, the No Child Left Behind Improvements Act of 2005. The bill contains over 40 provisions that would improve the implementation of the current federal law. In June, 2006, Representative Don Young (R-AK) introduced H.R. 5709, the No Child Left Behind Improvements Act of 2006, which incorporated all of the NSBA recommendations. Co-sponsors of H.R. 5709 included Representatives Steven R. Rothman (D-NJ-9), Rob Bishop (R-UT-1), Todd Platts (R-PA-19), and Jo Bonner (R-AL-1). In January 2007, Rep. Young reintroduced his bill as the No Child Left Behind Act of 2007, H.R. 648. The bill's cosponsors to date include Representatives Charlie Melancon (D-LA-3), Steven Rothman (D-NJ-9), Jo Bonner (R-AL-1), Thaddeus McCotter (R-MI-11), and Todd

Platts (R-PA-19), verifying strong bi-partisan support for these important improvements to the current law. This comprehensive bill addresses the key concerns of local school boards, including those provisions related to the accountability and the adequate yearly progress (AYP) framework. This bill would:

Increase the flexibility for states to measure adequate yearly progress (AYP), including growth models.

Grant more flexibility in establishing goals and determining AYP targets.

Create a student testing participation range, providing flexibility for uncontrollable variations in student attendance.

Allow schools to target resources to those student populations who need the most attention by applying sanctions only when the same student group fails to make adequate yearly progress (AYP) in the same subject for two consecutive years.

Ensure that students are counted properly in AYP reporting systems.

NSBA encourages you to review the No Child Left Behind Improvements Act of 2007, H.R. 648 in its entirety. However, for your convenience we have enclosed a copy of our Quick Reference Guide to the bill that provides the recommended provisions and a brief rationale.

NSBA very much appreciates the opportunity to submit a written statement for the Record, and we look forward to working closely with you and your staffs to complete the reauthorization process during this First Session of the 110th Congress. We will also provide you with recommended legislative language which should be helpful to your staff in drafting the new bill.

Questions concerning our specific recommendations may be directed to Reginald M. Felton, director of federal relations.

Sincerely,

MICHAEL A. RESNICK,
Associate Executive Director.

Chairman MILLER. And, with that, the committee will stand adjourned. And, again, thank you so very much.

[Whereupon, at 12:51 p.m., the committee was adjourned.]

○