

Testimony concerning the measurement of student achievement growth

Allan Olson
Co-Founder and Chief Academic Officer
Northwest Evaluation Association
March 21, 2007

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.

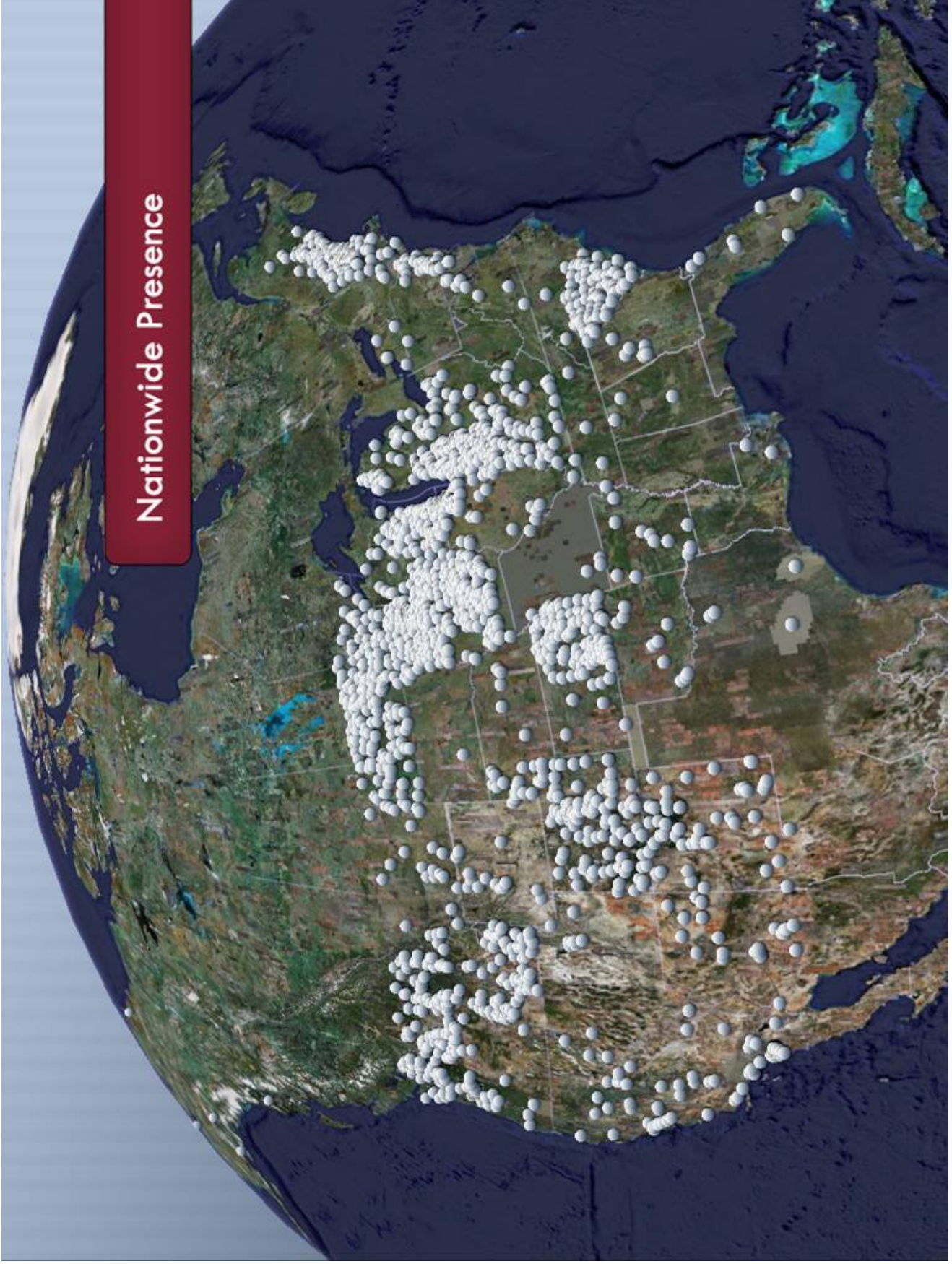
After all, the central issue is how we help the current generation of students meet our expectations. Measuring growth of each child gives us information that we can use to improve the growth of all of our students. At the same time, information about growth at the class and school level helps us describe 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.

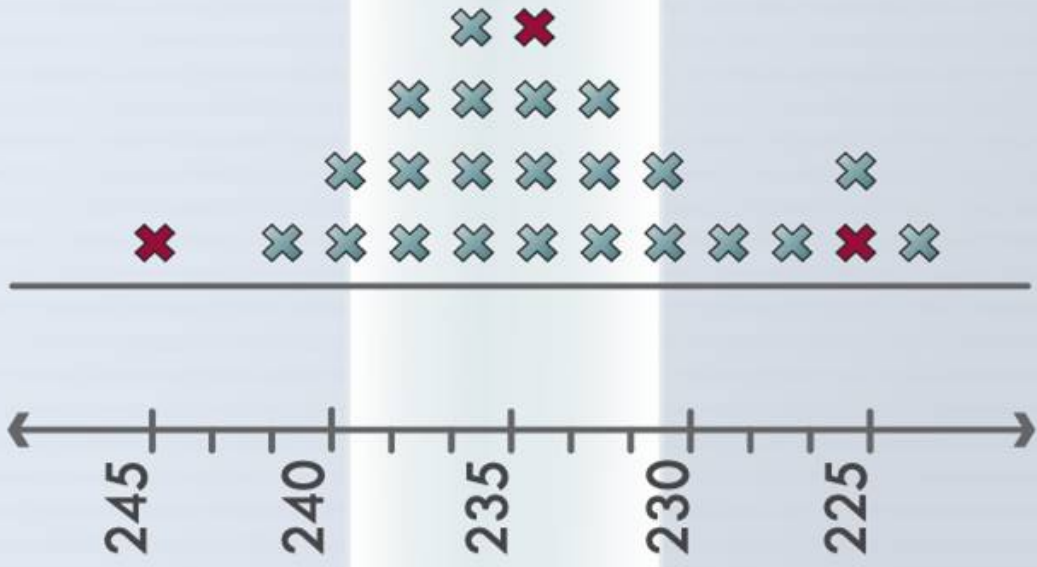
Thank you for the opportunity to share our experience and data with you this morning.

Nationwide Presence





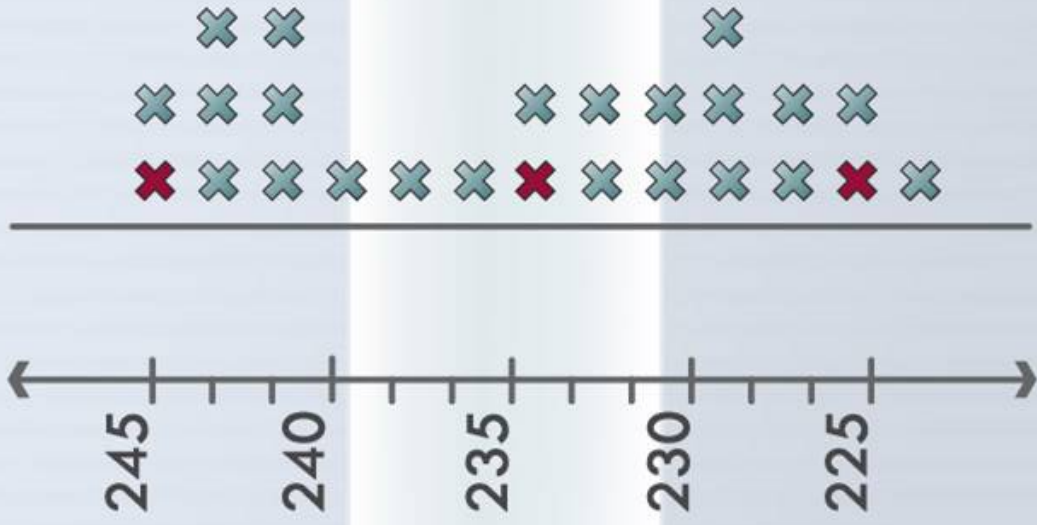
Typical Class Distribution



We typically think of a classroom and grade as having a bell shaped distribution with the bulk of kids performing within the area tested by a grade limited assessment.



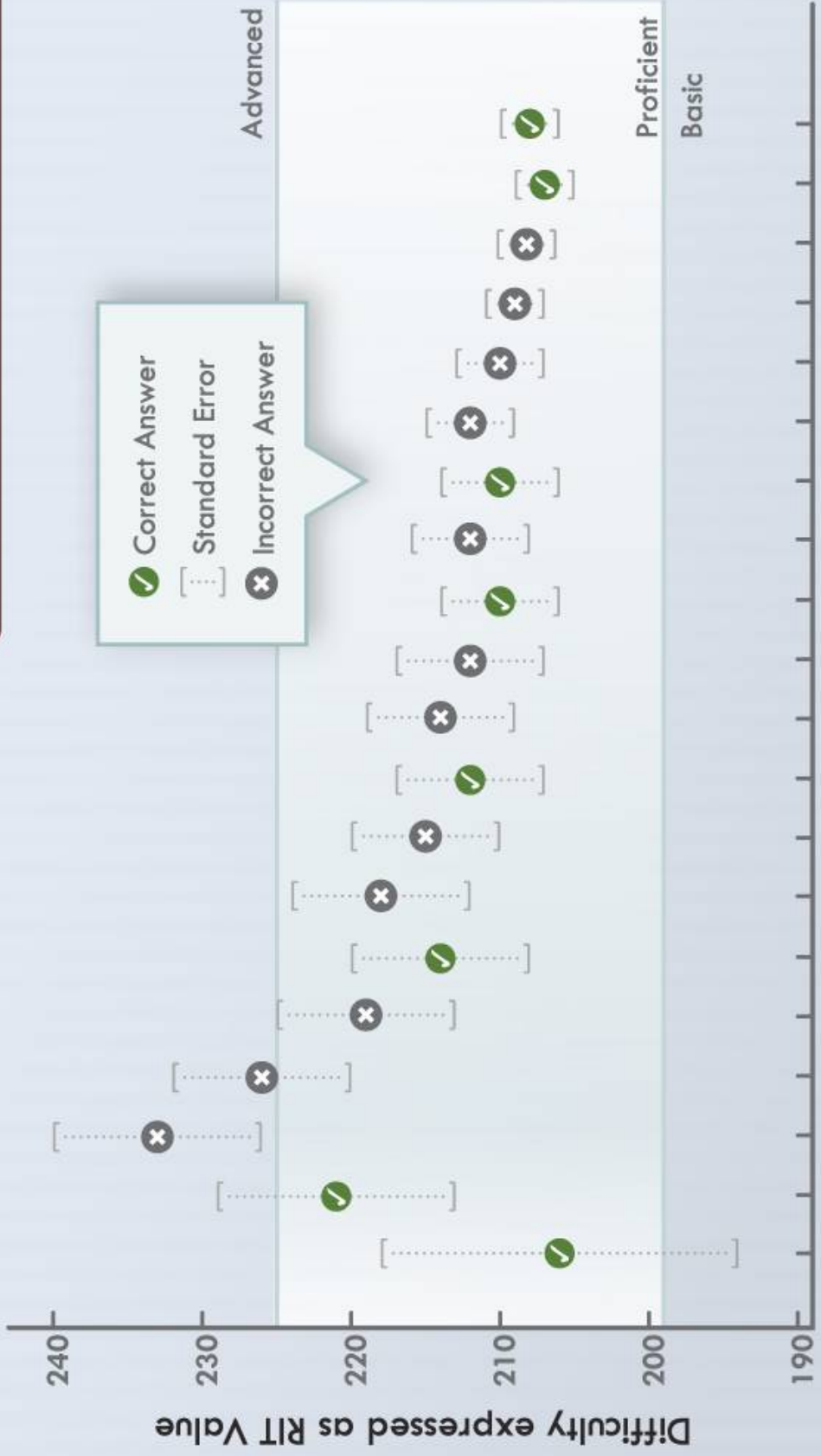
Typical Class Distribution



What if the bulk of kids were no longer represented by the middle?

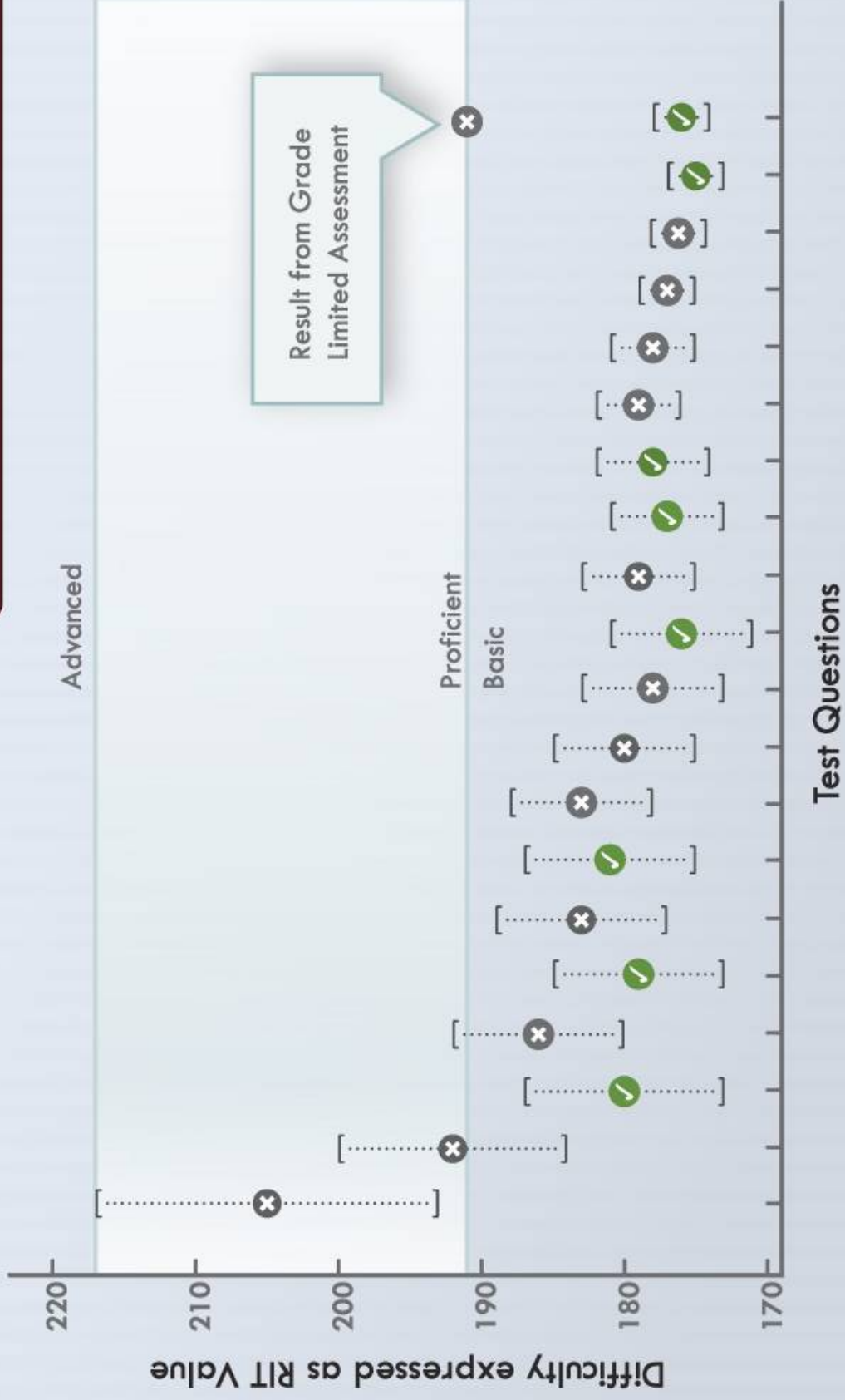


Adaptive Test Progression



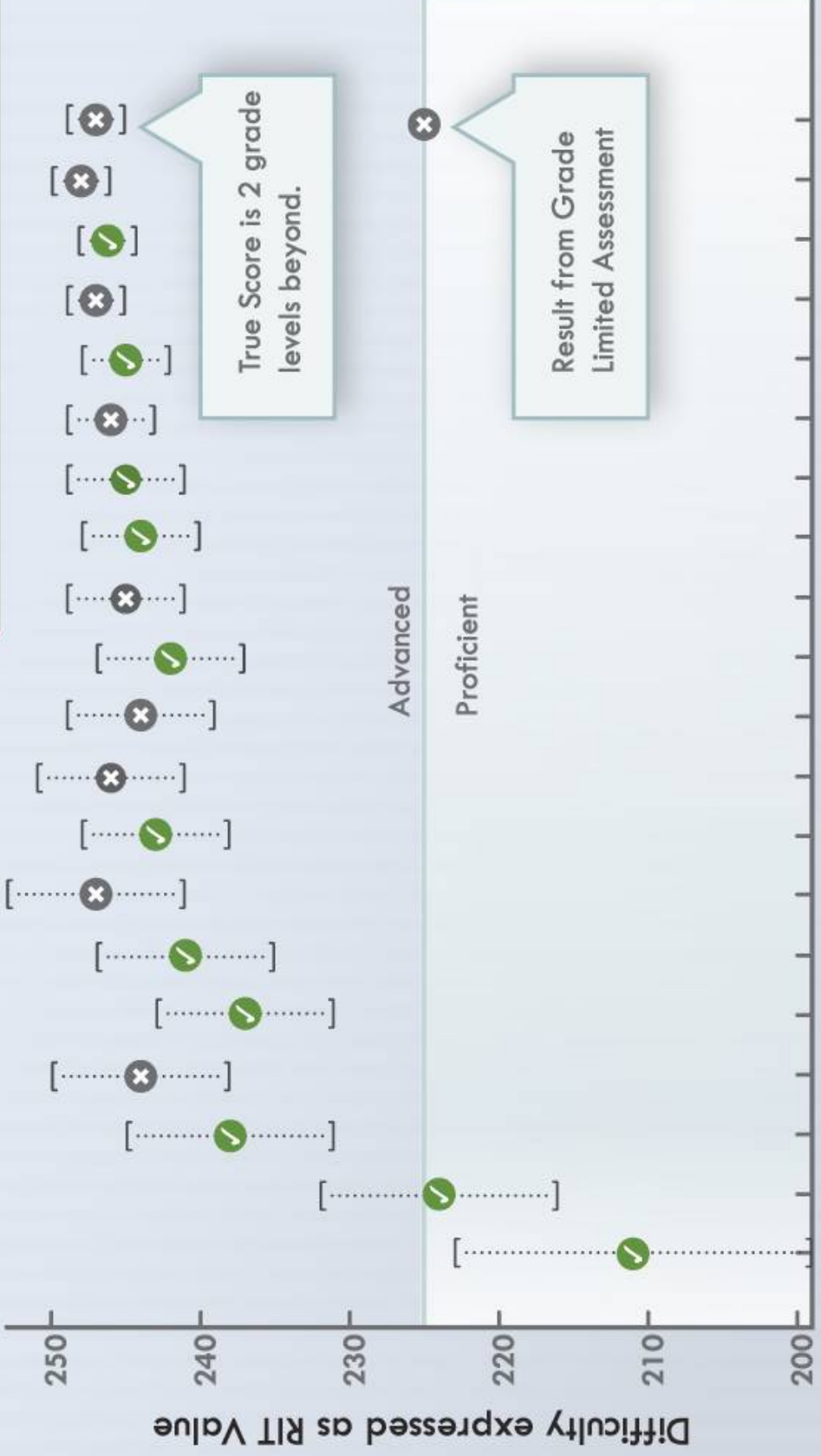


Adaptive Test Progression





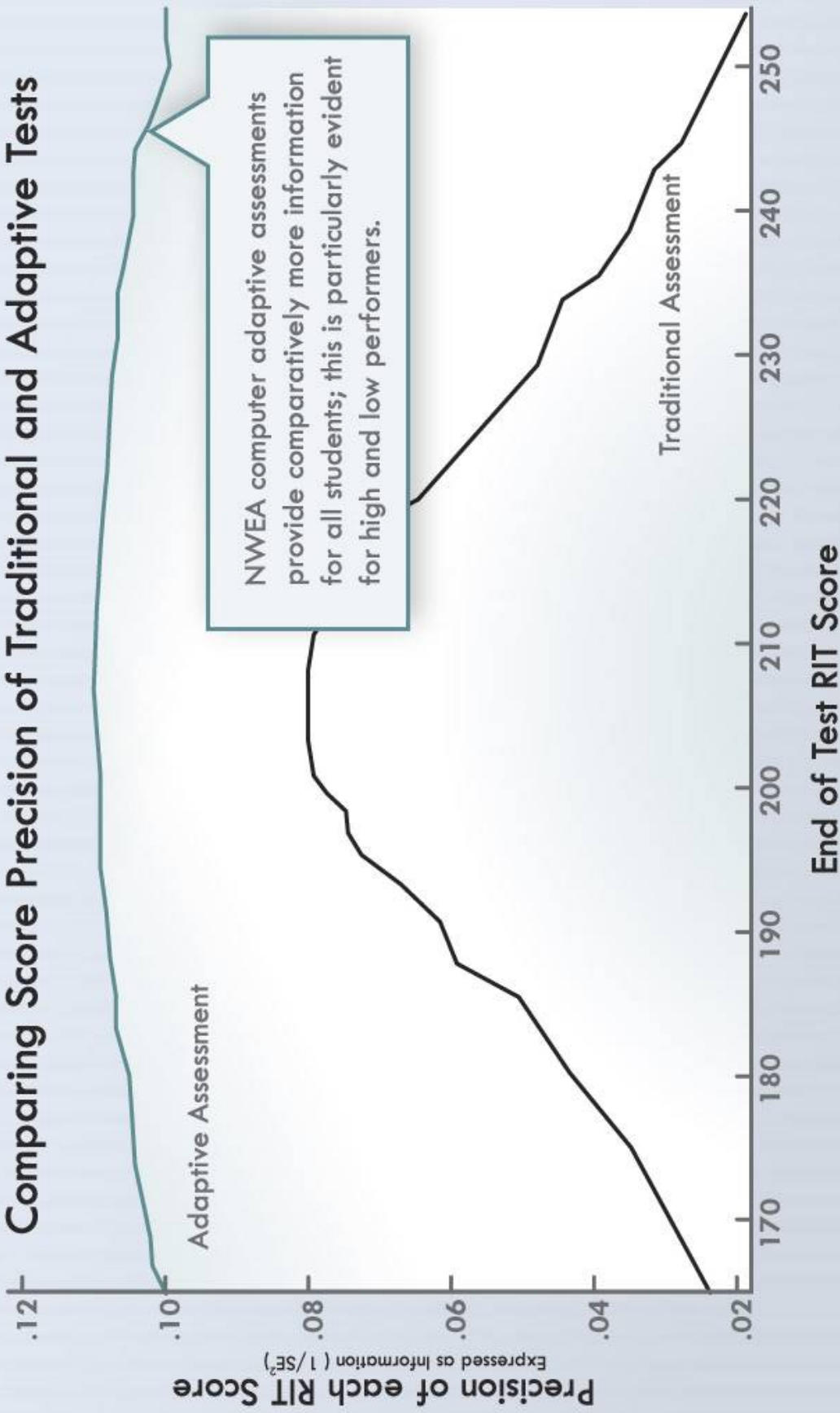
Adaptive Test Progression



Test Questions

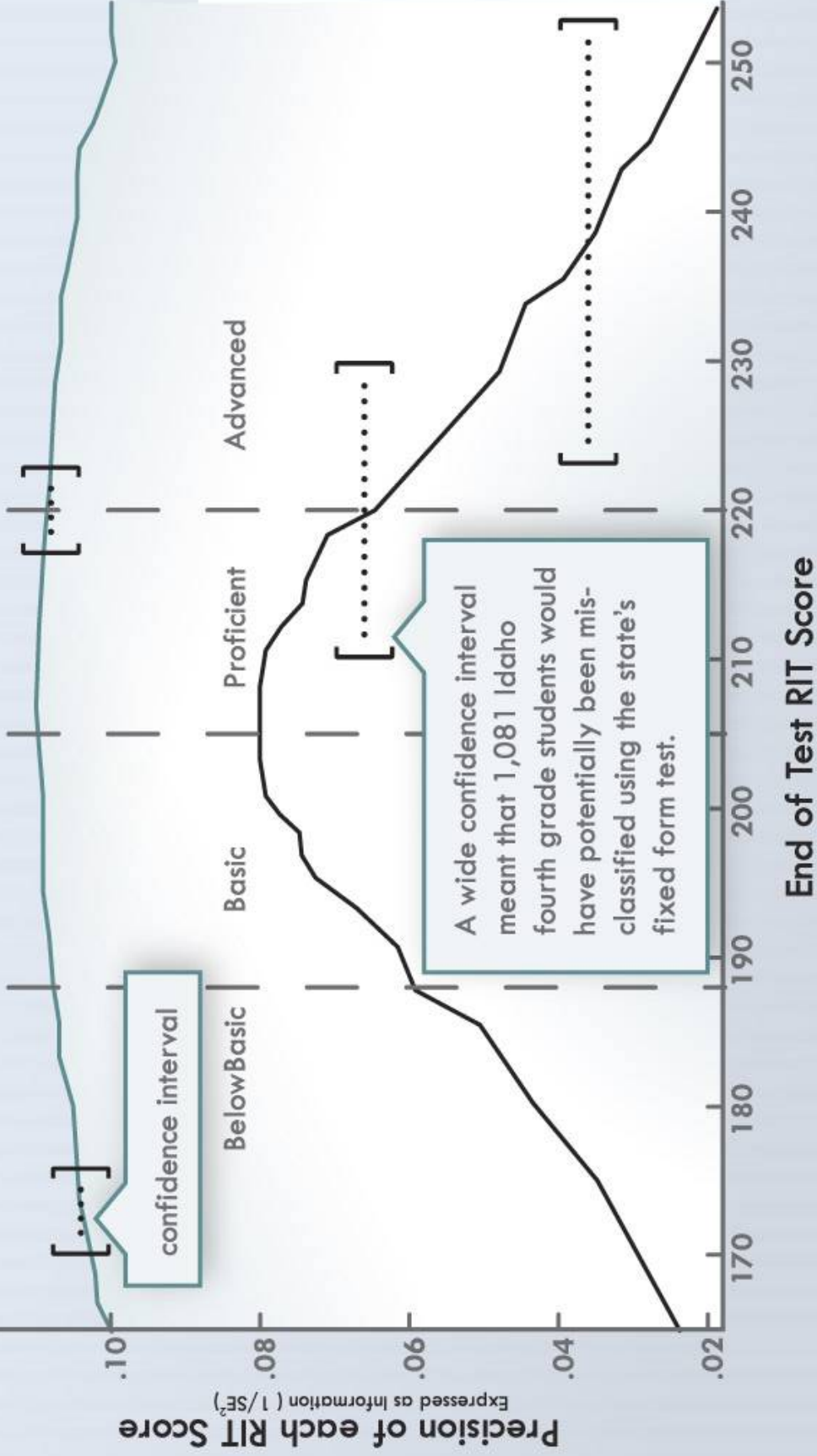


Comparing Score Precision of Traditional and Adaptive Tests





Comparing Score Precision of Traditional and Adaptive Tests



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 improvements, 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 know 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> or contact Matt Chapman at 503-624-1951.