NWEA™

Assessments with Integrity

How assessment can inform powerful instruction

A comprehensive ebook to help you use assessments with integrity to measure student growth.

January 2015

Introduction

After devoting nearly 40 years to researching and developing high quality assessments, we at Northwest Evaluation Association™ (NWEA™) have discovered one thing: assessment, when done well, can make a profound contribution to helping all kids learn. But for any test to make a real difference for student learning, it must be built with integrity.

Testing for the sake of testing helps no one; it is a waste of time for administrators, teachers, and students. On the other hand, assessing students with a rigorous, research-based tool can yield insights that help teachers deliver powerful, differentiated instruction to every student at the right time—allowing each individual to learn and grow.

A truly useful assessment must begin by meeting students where they are. In addition, crucial assessment components—such as norms, standard error of measure, scales, and item pool depth—must be built with integrity in order to yield actionable data that helps educators deliver the right instruction for each student.

Assessments developed with integrity are powerful tools that support educators and students in learning. NWEA delivers assessments educators can rely on to provide the powerful data they need to help every student grow.

Contents

Meeting Students Where They Are

Want to help a student grow? Find out where to start.

Jean Shields Fleming, Director of Communications, NWEA

Education is often described as the great equalizer. It is a force that can open opportunities for all students, regardless of their socioeconomic circumstances. It can fulfill the American promise that personal initiative can open new worlds. Yet students in today's schools are hardly uniform, and many have circumstances that put them at a disadvantage. This is especially true for students from fragile populations. A few statistics paint a picture of the disparity:

- 22% of all children in the U.S.—some 16 million kids—live at or below the poverty line.[1]
- Nearly 1.2 million students were homeless in 2012, 85% more than were reported prior to the recession that began in 2008.[2]
- The dropout rate for students living in poverty is 4.5 times greater than for students in higher income brackets.[3]
- In 2013, nearly 60% of all fourth- and eighth-graders were considered "not proficient" in reading and mathematics as measured by their end-of-year state assessments.[4]

It's clear that not only are economic gaps between children real and growing, but that those gaps have a profound impact on academic performance. To deliver on the promise of education as an equalizing force, America's teachers and administrators need to know where their students are starting their school journeys—to meet each student wherever he or she is—and have the tools to measure growth along the way.

Assessment Results Help Focus Efforts

A school system has many kinds of questions to answer. Teachers need to know where their particular students are starting and how they are growing toward goals—plus instructional information to move the student forward and differentiate instruction so all students learn. Principals need to understand how each class is performing and how the school as a whole is tracking toward established benchmarks. District administrators want to see overall trends and make sure the district is on track to meet accountability requirements.

No single assessment can meet all of these purposes—nor should it. Using multiple measures allows educators to cross-check their data and answer different educational questions with the appropriate tools. But when it comes to driving individual learning, especially for fragile populations, formative and interim assessments have a critical role to play in providing the information educators need to close achievement gaps. To understand where all students are on their learning path, an adaptive assessment can be an invaluable tool, provided it meets certain criteria: measuring growth regardless of grade and gathering data efficiently.

Beyond grade measurement

To understand the disparities among students—to measure the gap—the assessment must be able to measure students who are performing on, above, or below grade level. There is a place for understanding grade level proficiency (in fact, federal accountability frameworks demand it), but to actually teach each student as he or she is, today, the teacher needs to know where the starting line is.

Adaptive tests, which adjust with each test question, provide the clearest picture of that starting line.

Many tests adapt only after several items have been presented, which does not return the same precision as a test that adjusts in real time in reaction to every single student response. In addition to this true adaptivity, the test also needs a deep pool of items to draw from in order to ensure that students are seeing new questions with appropriate depth of knowledge each time they take a test. And of course, an assessment must use a stable scale, which is the only way to accurately show a student's growth over time, regardless of grade level performance.

Assessment efficiency

Efficient assessment is also crucial to meeting students where they are because it returns actionable data without sacrificing instructional time. Efficiency also refers to obtaining quality and precise data from the assessment instance. Do not sacrifice accurate data with shorter tests, as they could give incorrect information about what students know and are ready to learn. Adaptive tests—like Measures of Academic Progress® (MAP®) interim assessments—can pinpoint student growth and instructional needs accurately in a relatively short amount of time. MAP assessments take about an hour, and give students and educators information they can immediately use to move learning forward.

- Students immediately see how they scored on the test.
- Teachers see how the class is performing and can use this information to set goals with students, create flexible groups to differentiate instruction, and communicate with parents.
- Principals get a view of their entire school and can direct resources to meet specific needs.
- District administrators can see how each school is performing and make adjustments based on reliable information.

In an era of tight budgets, and large and varied classes, this efficiency brings exceptional value. When teachers have high-quality information about each student's actual learning needs, regardless of grade placement, they can make heroic growth possible for all kids.

Jean Shields Fleming brings over 25 years of experience in education to her role at NWEA. She began as a middle school reading teacher in the Berkeley, California public schools. There, she developed a curriculum focused on engaging students in career explorations to foster a love of reading. She served as lead instructional designer for an online reading curriculum, held senior editorial positions with Technology & Learning magazine and Scholastic.com, and managed global communications for the Intel Foundation's professional development program.

The Need for Norms

Quality norms help students and teachers achieve more

John Wood, Senior Analyst, NWEA

Teachers do a better job of reaching each student when they don't teach in a vacuum—and that's where norms come in. An educational norm is simply a picture of the typical level of performance for any given group of students based on characteristics such as age range, grade level, or geographic area. MAP and MAP for Primary Grades (MPG) assessments provide educators with high-quality norms that help educators see their students' learning in a wider context.

Types of Norms and How Educators Use Them

Norms help educators see if a student is growing at an expected pace, regardless of where the student started. NWEA provides both status and growth norms that allow educators to compare students' academic performance to peers. Often, teachers use norms to help explain to parents and students what a given assessment score means, and they often use norms to help make important decisions about placement in both achievement and Response to Intervention (RTI) programs. Norms may be used by schools for analysis and evaluation of performance and programs. In all cases, having accurate norms drawn from appropriate sample sizes is crucial for educators.

Obtaining Quality Norms

As with any statistical model, norms are most useful when they accurately reflect the population they represent and when they allow for closer analysis of various sub-groups. Because of the sample sizes and the advanced statistical techniques used to develop NWEA norms, educators can rely on the validity of the data to make important decisions about students, schools, and programs. NWEA mines five terms of test records to create large, nationally representative samples—over 20,000 students per grade. We create our data sets to ensure we represent United States public school demographic groups and geographic regions appropriately. Further, NWEA creates norms that reflect various weeks of instruction during the school year rather than simply taking a single snapshot. You can see how NWEA accounts for weeks of instruction in the report, which shows means for fall, winter, and spring for each grade. This feature makes NWEA norms unique.

Status Norms

NWEA provides achievement, or status, norms for each grade in math, reading, language usage, and science. Teachers can view these norms based on any number of instructional weeks through the year to gain a more precise estimate of a student's current achievement relative to his or her peers. Status norms, presented as a percentile rank, can be useful for placing students into various programs. For example, MAP and MPG scores have met the standards of the National Center for Response to Intervention (NCRTI) to be used as a universal screener for placement in response to intervention (RTI) programs. NWEA also provides school-level norms by subject and grade that parallel the student norms.

Growth Norms

The real power of MAP and MPG assessments is in measuring growth over time and comparing each student's measured growth to growth norms. NWEA makes this easy for educators with the Conditional Growth Index (CGI), which allows for comparisons based on subject, grade level, instructional weeks, and students' starting RIT score (the score given by the MAP test). Using growth norms, teachers can understand not only that a student grew, but how much that student grew in comparison to other students who had similar starting RIT scores and weeks of instruction. Accounting for instructional progress in growth norms is something only NWEA offers. Growth norms provide teachers the necessary context for setting individual student growth targets, a powerful way to help students take ownership of their own learning. At the school level, growth norms express the progress of groups of students and can be useful for making inferences about programs and instructional approaches.

Norms and Instructional Content

Over the last couple of years, NWEA has leveraged its norms data to help teachers differentiate instruction by linking our data to content in both open educational resources like Khan Academy and other instructional content that schools have purchased. These links can create individualized learning paths for each student by suggesting content that is in the student's zone of proximal development. Equally important, these individualized learning paths can help parents and guardians support students at home.

John Wood is Senior Analyst for Assessment and Education on the Academic Services team at NWEA. He has worked extensively with the Common Core State Standards (CCSS) since they were issued in draft form, and frequently gives presentations to NWEA partners on the relationship of MAP to the CCSS. He provides subject matter expertise to many teams at NWEA, as well as to partners. He is constantly working on new ways to connect MAP data to instructional content in order to help educators create assessment-informed learning paths for each student.

Making Sense of Standard Error of Measurement

Nate Jensen, Ph.D., Research Scientist, NWEA

Imagine for a moment that you're starting a new exercise and diet program to lose ten pounds in the next year. In order to track your weight loss progress, you decide to purchase a new scale. The store, however, only carries two different styles of scales—the first is much cheaper, but only guarantees its estimates of weight within ± 5 pounds (a range of 10 pounds), whereas the second one, while more expensive, guarantees its estimates to within ± 0.1 pounds (a range of 0.2 pounds). Which scale would you buy?

The answer seems obvious. It wouldn't be very helpful if every time you stepped on the scale you couldn't be certain if your weight was actually five pounds heavier (bummer), five pounds lighter (victory!), or somewhere in between. In order to track your progress over the weeks and months, it would be important to have a scale that would return an accurate estimate of what you actually weigh every time you step on the scale.

This same principle holds with student test scores. If you want to track student progress over time, it's critical to use an assessment that provides you with accurate estimates of student achievement—assessments with a high level of precision. When we refer to measures of precision, we are referencing something known as the Standard Error of Measurement (SEM).

Before we define SEM, it's important to remember that all test scores are estimates of a student's true score. That is, irrespective of the test being used, all observed scores include some measurement error, so we can never really know a student's actual achievement level (his or her true score). But we can estimate the range in which we think a student's true score likely falls. And, in general, the smaller the range, the greater the precision of the assessment.

SEM, put in simple terms, is a measure of precision of the assessment—the smaller the SEM, the more precise the measurement capacity of the instrument. Consequently, smaller standard errors translate to more sensitive measurements of student progress.

On MAP assessments, student RIT scores are always reported with an associated SEM, with the SEM often presented as a range of scores around a student's observed RIT score. On some reports, it looks something like this:

Student Score Range: 185-188-191

So what information does this range of scores provide? First, the middle number tells us that a RIT score of 188 is the best estimate of this student's current achievement level. It also tells us that the SEM associated with this student's score is approximately 3 RIT—this is why the range around the student's RIT score extends from 185 (188 - 3) to 191 (188 + 3). A SEM of 3 RIT points is consistent with typical SEMs on the MAP tests (which tend to be approximately 3 RIT for all students).

The observed score and its associated SEM can be used to construct a "confidence interval" to any desired degree of certainty. For example, a range of ± 1 SEM around the observed score (which, in the case above, was a range from 185 to 191) is the range within which there is a 68% chance that a student's true score lies, with 188 representing the most likely estimate of this student's score. Intuitively, if we specified a larger range around the observed score—for example, ± 2 SEM, or approximately ± 6 RIT—we would be much more confident that the range encompassed the student's true score, as this range corresponds to a 95% confidence interval.

So, to this point we've learned that smaller SEMs are related to greater precision in the estimation of student achievement, and, conversely, that the larger the SEM, the less sensitive is our ability to detect changes in student achievement. But why is this fact important to educators?

If we want to measure the improvement of students over time, it's important that the assessment used be designed with this intent in mind. And in order to do this, the assessment must measure all kids with similar precision, whether they are on, above, or below grade level. Recall, a larger SEM means less precision and less capacity to accurately measure change over time, so if SEMs are larger for high- and low-performing students, this means those scores are going to be far less informative, especially when compared to those students who are on grade level. Educators should consider the magnitude of SEMs for students across the achievement distribution to ensure that the information they are using to make educational decisions is highly accurate for all students, regardless of their achievement level.

An example of how SEMs increase in magnitude for students above or below grade level is shown in the following figure, with the size of the SEMs on an older version of the Florida 5th grade reading test plotted on the vertical axis relative to student scale scores on the horizontal axis.[5] What is apparent from this figure is that there is a tremendous amount of imprecision associated with the test scores for low- and high-achieving students. In this example, the SEMs for students on or near grade level (scale scores of approximately 300) are between 10 to 15 points, but increase significantly for students the further away they get from grade level. This pattern is fairly common on fixed-form assessments, with the end result being that it is very difficult to measure changes in performance for those students at the low and high end of the achievement distribution. Put simply, this high amount of imprecision will limit the ability of educators to say with any certainty what the achievement level for these students actually is and how their performance has changed over time.

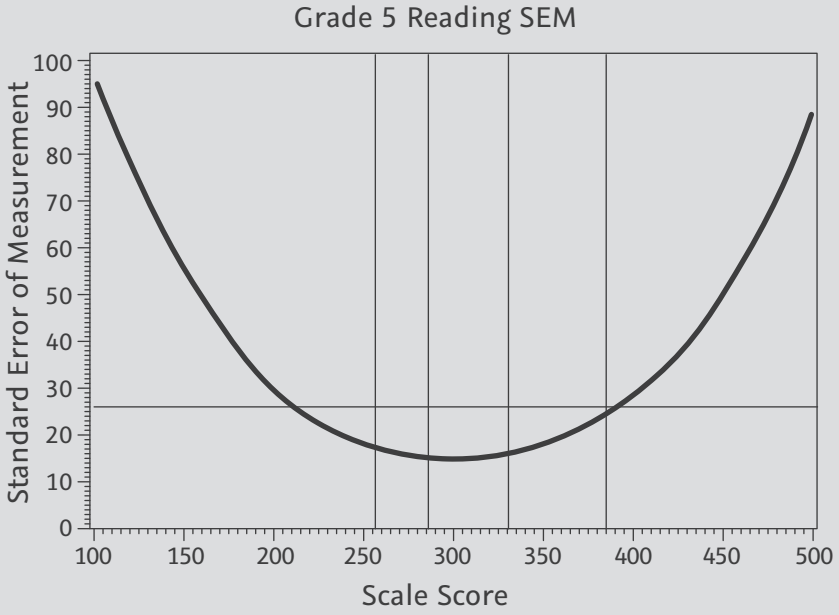


Figure 1: Standard Errors of Measurement by Student Achievement Level, 5th Grade Reading, Florida Comprehensive Assessment Test (2006)[5]

Of course, the standard error of measurement isn't the only factor that impacts the accuracy of the test. Accuracy is also impacted by the quality of testing conditions and the energy and motivation that students bring to a test. In fact, an unexpectedly low test score is more likely to be caused by poor conditions or low student motivation than to be explained by a problem with the testing instrument. To ensure an accurate estimate of student achievement, it's important to use a sound assessment, administer assessments under conditions conducive to high test performance, and have students ready and motivated to perform.

In short, just as you would want an accurate scale if you were tracking your progress in an exercise and diet program, it is critical that the assessment used in your school to track student progress provides achievement estimates for all students that are as precise as possible. And, as we have discussed in this paper, the key metric on which to focus to determine if that is the case is the standard error of measurement. The SEM of a test score isn't always the easiest statistical term to understand. But, if you remember that smaller SEMs lead to more accurate estimates of student achievement and growth, you will be better able to interpret and evaluate the performance of your students.

Prior to joining NWEA three years ago, Nate Jensen put his skills to use as a teacher and as a senior research associate at the office for education policy at the University of Arkansas. Over a decade of experience in the education field has given him a profound appreciation of the importance of data in the classroom. His passion lies in helping teachers and school leaders understand how to use metrics to deeply understand what each student needs in order to grow.

Measuring Growth with a Stable Scale

A strong, stable scale is key to accurately measuring student growth

Nicole A. Zdeb, Senior Manager, Academic Services, NWEA

At NWEA, we are passionate about growth and what it means in the educational journeys of the students we serve. Growth means that learning is happening. Students are acquiring the concepts and skills they need to flourish in the world and be the authors of their own stories. Measuring that growth is crucial, and the stability of a measurement scale over time is necessary to measure growth accurately. But what is a scale and what does stability mean?

Measuring Latent Traits

Scale is one of those words that befuddles learners of English because there are various and unrelated meanings, from the covering on certain animals, to a cause of blindness, to an instrument with graduated degrees. When we talk about scales in educational measurement, we're talking about something akin to the latter definition: a construct that indicates the degree of student ability in a certain area, such as mathematical reasoning. We call this ability a latent trait because we can't see the amount of a student's mathematical reasoning ability directly, like we can see a student's hair color or the shoes she is wearing. Even though we can't physically interact with it, we know mathematical reasoning is a real thing. It is similar to a psychological state such as happiness—it exists in the interiority of a person. Quantifying, or measuring, latent traits is a subtle endeavor to which we bring to bear the power of statistical modeling.

To measure a latent trait, we must first elicit it. We elicit evidence of a latent trait through the use of instruments such as test items, performance tasks, and writing samples. The evidence is a proxy, or representative, of the trait itself. We infer that a person is happy by certain indicators such as body language, laughter, or smiling. Likewise, we infer the degree of a latent trait such as mathematical reasoning by the evidence we get from eliciting that trait.

Stable, Equal Interval Scale

Once we elicit evidence of a latent trait, we measure that evidence by using a scale. The scale used by MAP is an equal interval measurement called the RIT scale. An equal interval scale provides a specific kind of information about order, namely that there is the same distance between points on the scale, or the same amount of underlying quality. Another example of an equal interval scale is a thermometer. The value of an equal interval scale is that it is consistent and objective; such qualities help make a scale stable over time.

Scale stability means that scales maintain their measurement characteristics, allowing for comparisons of assessment scores among groups of students, growth estimates, and longitudinal studies. For example, a RIT score of 215 in 1975 would be equivalent to a RIT of 215 in 1995. Maintaining this stability of scale—so that our scale retains its measurement properties in exactly the same way year after year—is a crucial part of what we do at NWEA.

Guarding Against Scale Drift

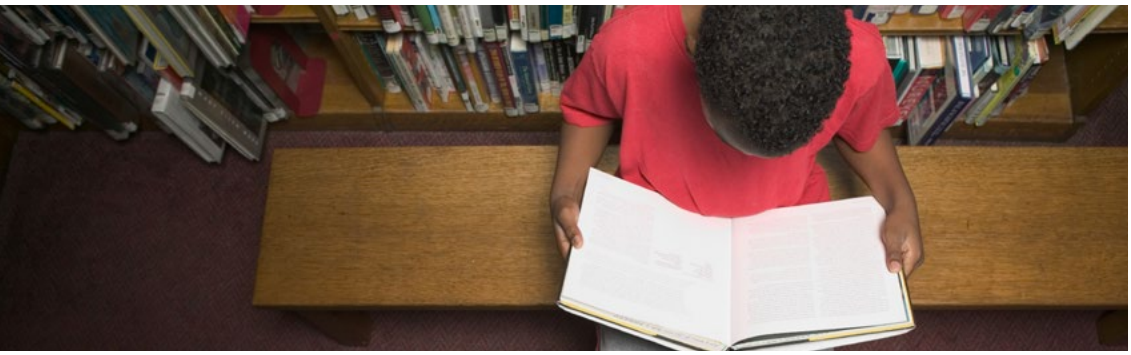
Not all educational measurement scales are inherently stable. Many scales can and do drift over time, meaning that the underlying quantity of what is

being measured shifts. Item calibrations can become more difficult or less difficult as time goes on. This is disastrous for measuring growth or making any sort of longitudinal comparisons. There are a number of reasons a scale might drift over time:

- Curricular changes (including pedagogical approaches)
- Changes in standards or what is valued in a content area
- Changes in testing populations over time
- Changes in the stakes of a test or assessment, its social context, or the meanings and consequences attached to it
- Changes in the intended purposes of the assessment

Guarding our scale's integrity in the face of such changes allows NWEA to maintain a scale with remarkable stability over time. This stability gives us confidence in the educational decisions that are made as a result of the data gleaned from our MAP assessments. We are aware of the importance of maintaining scale stability, so we make it a priority to monitor our scales for drift. The scale connects it all—items to scores to students. We ensure our data are the gold standard in measuring growth because we know that every student can grow. Reporting that growth accurately makes a difference in the lives of students across the country and world.

Making a difference in the lives of children, teachers, and families is what has driven Nicole Zdeb's work at NWEA over the past decade. As a specialist in content and curriculum, she's brought her prodigious experience as an educator to bear on the Academic Services department of NWEA. Outside the office, Nicole devotes her time to supporting student literacy, and has had poems published in numerous literary journals.



How Deep Is the Pool—and Why Does it Matter?

How item pool depth affects your assessment's accuracy

April Roethel, Senior Analyst, Assessment and Education, NWEA

To dive safely into a swimming pool, the water must be deep enough to extend past your head and your feet. Similarly, to ensure the validity and effectiveness of computer adaptive tests (CATs), the item pool must be deep enough to stretch above and below a student's entry point.

An essential part of a CAT is a well-constructed item pool. One crucial requirement of a well-designed item pool is that it includes enough items to enable the building of numerous individualized tests that align to students' varying ability levels—this is what we mean by the depth of the pool. The ideal item pool should also include enough breadth to cover the scope of the content domain.

Why is item pool depth so important in computer adaptive assessments?

Unlike paper-and-pencil tests that are fixed in content, CATs adapt to individual student performance. They get harder or easier depending on how a student is performing on the test, which requires a deep item pool from which many different tests can be drawn.

A student's grade level is not necessarily his or her instructional readiness point; this is why a CAT must adapt to measure on-, above-, and below-grade abilities. An assessment that informs educators about each student's instructional readiness draws on content that spans across grades. A deep item pool can provide this because it will be stocked with items that correspond to many different grade levels.

How many items are enough to assess a skill?

A common question that comes up when evaluating an assessment tool is, "How many items should it include?"

The appropriate size of the item pool depends on four main factors.

Precision is the first factor to consider, as it relates to the "estimate of student achievement that is desired."[6] The more precision you desire, the larger your item pool needs to be. If you are aiming to get just a rough estimate, you can use a smaller item pool.

- Range is another significant factor. How broad or narrow is the range of achievement to be measured? A larger item pool will be required for assessment that is very broad, since it will include items with a large range of difficulty. For example, if an assessment is being used to measure students' performance at multiple depth of knowledge (DOK) levels, it will require a greater range of items than an assessment concerned with only one DOK level.
- Stakes are a third factor that will determine the item pool size requirement. If a CAT is very high stakes, students might be more likely to game the test. Large item pools improve the chance that examinees receive a different set of test items for every test administration, making it impossible to cheat the system.
- Number of times a CAT is administered is a fourth factor of importance. If an assessment is administered to the same students multiple

times a year, for instance, the item pool must be large enough to ensure that a student doesn't see any item more than once.

The goal is to have a sufficient number of items in each desired content area to assemble an individual test with the balanced content coverage required by the test.[7]

The importance of field testing and calibration

A deep pool of items isn't very valuable if the items themselves aren't high quality. Field testing enables identification of items that are performing atypically. Poorly performing items should be removed from the item pool as soon as they are identified to avoid proficiency estimation errors. Additionally, a rigorous calibration process builds confidence that an item is likely a good measure of the attribute in question. This is another instance where a deep pool of samples creates a high degree of accuracy. NWEA bases calibration on more than 1,000 student responses; this is one of the most stringent calibration processes in the education assessment field.

Why MAP?

The adaptive nature of MAP assessments allows them to measure each student with a high degree of precision, regardless of whether the student is performing on, above, or below grade level. MAP is a well-constructed interim assessment with a deep item pool. This means that it has enough items to assess depth of knowledge and true performance of all students. This deep item pool also allows MAP to control for item exposure, providing students a unique testing experience each time, with no opportunity to memorize items or item order.

MAP items undergo a rigorous development calibration process, too. High standards in item development and calibration contribute to a robust item pool. And this deep item pool is still growing, with new item types and functionalities being added continually to better assess higher-order thinking skills and the full range of each content domain. The item pool is dynamic and responsive to what kids are learning in the classroom.

Having taught in the classroom for 15 years, April Roethel knows first-hand how important it is to help all kids learn. She's spent 30 years in the education field developing resources to support that mission. As both a Senior Analyst and a Senior Content Specialist at NWEA, April has relied on her experience and passion to help develop assessments that truly support teachers and students in the classroom.

Measuring Growth Right

Accurate growth measurement leads the way to proficiency for every student

Nicole A. Zdeb, Senior Manager, Academic Services, NWEA

For nearly 40 years, we at NWEA have understood the power of student growth and have dedicated ourselves to measuring it better than anyone else. Growth can occur regardless of where students come from and what they might have working against them. Growth is possible for all students, and it is a meaningful indicator of learning. Measuring student academic growth is how we can track each student's progress toward proficiency and achievement—and continually adjust course to support learning.

Growth is a vital gauge of student progress, so it's crucial to measure it accurately. Measuring growth is a challenge that NWEA is well-equipped to meet—and has for millions of kids, in thousands of classrooms, across the world. To measure growth well and provide actionable data for all students regardless of their starting points, an assessment must do three things:

- Have a stable scale as a foundation. A stable scale results in data that are comparable over time for individual students and groups of students. Growth, by its very nature, can only be measured over time, in snapshots. A stable scale allows for growth to be measured accurately.
- Include cross-grade content. Growth is possible for every student. But to measure growth for each individual, the assessment has to find each student where she or he is—which may be on, above, or below grade level. Assessments that only include on-grade level content can't serve the students who are challenged or gifted. It can only tell the growth story of a fraction of the classroom.
- Provide context for the assessment results. A score with no context has very little utility—it exists in the abstract. Context, such as understanding what that score means in relation to other students—in the classroom, in the grade, and across the nation—helps an educator make informed instructional decisions. One type of context is comparing a student's score to a pre-set benchmark. This helps educators understand if the student is achieving at an expected level or needs more support. Another important sort of context is provided by growth norms and growth targets. These help an educator understand if the student is growing on par with her peers and whether she is projected to meet achievement standards. If she isn't, this sort of context allows educators to make adjustments in time to get the student on track for success.

NWEA creates assessments that meet all these criteria for measuring growth right. MAP assessments have a demonstrably stable scale, include high-quality cross-grade content, and provide rich contextualization of assessment data. We invest in measuring growth accurately and are committed to doing so for every student. Academic growth is a beautiful thing—it reflects evolution, expansion, and empowerment. Proficiency for all students is the destination, and it's growth that will get us there.

Making a difference in the lives of children, teachers, and families is what has driven Nicole Zdeb's work at NWEA over the past decade. As a specialist in content and curriculum, she's brought her prodigious experience as an educator to bear on the Academic Services department of NWEA. Outside the office, Nicole devotes her time to supporting student literacy, and has had poems published in numerous literary journals.

Data That Are Actionable

Truly transformative assessments give educators data they can act on

Kristin Moran, M.Ed., Senior Curriculum Specialist, Professional Development, NWEA

Getting data into the hands of teachers and school leaders to inform instructional decisions is what gives assessment its power. Timeliness is key, as is structuring opportunities for application of the data. Using actionable assessment data can help all stakeholders answer important questions about student learning, such as:

- As a teacher, how can I adjust my instruction to meet students' needs? How will I know what kind of progress they're making?
- As a school principal, how can I ensure that the students in my school are tracking toward key milestones? How can I offer the best professional development to support teachers?
- As a district administrator, how can I evaluate our district's programs for improvement planning? What's working best, and what should we stop doing?
- As a parent, how do I know my child is receiving instruction that will extend her current knowledge and skills?
- As a student, what do I need to work on to reach my goals?

MAP assessments give students, parents, and educators more than just a score—these assessments deliver data that can actually be used in real time to make a difference in education. That's why it's so crucial that assessments provide actionable data. Data from MAP tests help educators differentiate instruction based on student readiness, set academic goals with students, and evaluate programs, including professional development programs.

Differentiate by Student Readiness

Carol Ann Tomlinson and Tonya R. Moon[8] describe differentiation as a "teacher's proactive response to learner needs." Such a proactive response allows a teacher to meet students within their zone of proximal development (ZPD), the optimal spot where instruction is most beneficial for each student—just beyond his or her current level of independent capability. The ZPD is not about a student's ability to learn, but rather about what skills and understandings the student is ready to develop with targeted assistance or scaffolding.

MAP assessments provide a grade-independent RIT score that measures academic growth, much like a yardstick measures physical growth. Teachers can align student RIT scores with normative data to analyze how their students compare to one another and the NWEA nationally normed group, revealing the diversity of skills within their classroom. Teachers can then target instructional content to students of varying skill levels by using the Learning Continuum, a tool that helps teachers pinpoint where individual students are ready to advance—and where they need help. The Learning Continuum translates student RIT scores into actionable learning statements that enable teachers to develop instructional plans for students. For instance, the Class View in the Learning Continuum groups students by RIT score bands to show where students fall on the RIT scale and what skills and concepts they are ready to learn. This informs decisions for flexible groupings, supporting differentiation based on student readiness.

Example: Differentiated Instruction in Practice

Ms. Ramirez is teaching a 4th grade lesson on the topic of perimeter. Using MAP data in the Class View, she determines three manageable flexible groups, the students associated with each group, and what learning statements are pertinent for each group.

RIT Range	Students	Learning Statements
171-200	J.A. Cambridge E.H. Horton L.L. Wojnarowski A.H. Frisino D.H. Engles	Determines perimeters of basic polygons with all sides labeled
201-210	J.L. Russell L.E. Kong J.B. Dotson D.A. Patel	Determines perimeters of basic polygons in which not all sides are labeled
211-220	R.N. Sandoval M.G. Moyer B.O. Esteban	Counts to find perimeters of complex figures

The learning statements support teachers, like Ms. Ramirez, in scaffolding instructional activities, effectively accessing each student's ZPD. Using content from the school's curriculum, along with MAP resources such as RIT to Resource and MAP to Khan Academy, teachers can build responsive lesson plans in alignment with learning statements. Pairing MAP data with formative assessment practices, teachers can adjust instruction in the moment based on student feedback.

Set Academic Goals

In their landmark book on classroom assessment, Jan Chappuis, Rick Stiggins, Steve Chappuis and Judith Arter[9] have shown that involving students in self assessment, goal setting, and reflecting on their learning is critical to quality classroom assessment. Teachers can use MAP growth projection data and learning statements to develop individual academic learning goals with students.

Example: Academic Goal-Setting in Practice

Mr. Waller reviews all of his 3rd grade class' growth projections for math. He notices Louis has a projected growth RIT score of 183, and based on his state's cut score ranges, Louis needs a score of 195 to meet proficiency. In a goal-setting conference, Mr. Waller and Louis dig deeper into the data using the Student Goal Setting Worksheet and uncover goal areas of strength and opportunity. Agreeing on a mid-year RIT goal of 190 and an end-of-year RIT goal of 195, they choose a few learning statements from the Learning Continuum to develop into specific student goals. Throughout the year, they track progress toward the goals.

Going beyond the individual student level, teachers and school leaders can identify strengths and areas for improvement in MAP goal performance areas for classrooms, schools, or entire districts. Creating a data-centric school culture, engaging students and parents in the goal-setting process, and celebrating student progress will help instill a culture of goal setting that has a lasting impact.

Evaluate Programs and Target Professional Development

Schools and districts can use data to evaluate curricula and intervention programs, inform changes in instructional practice, and target professional development. The use of MAP status and growth data can help identify what's working and point to successful programs that can be scaled up. It helps answer questions like: Did the students in our new math program experience higher rates of growth than other students? Where do our teachers need to focus instructionally? What kind of professional development will assist our district in targeting areas of concern?

Example: Program Evaluation and Professional Development in Practice

The Shaw School District implemented a new 4th grade reading program that emphasizes the College

and Career Readiness shifts in reading, including more focus on informational texts and eliciting evidence from texts. Analyzing MAP data from fall to spring revealed that the majority of individual students made significant gains, even though most of the students stayed within their achievement level (i.e. needs improvement, proficient, etc.). Instead of abandoning the reading program, the district decided to continue its use for 4th grade and extend it to the incoming 5th graders in order to chart its effects on longitudinal growth. Disaggregating the data presented specific goal areas that experienced lower rates of growth than other areas. The district targeted teacher professional development to address these areas in which students were exhibiting insufficient growth.

Make Actionable Data Accessible

Using student growth data to inform instruction can be a valuable and efficient tool for driving students' academic gains. To make sure your data can be used most effectively:

- Cultivate data literacy for all stakeholders, providing the tools and resources to understand critical issues, such as assessment types and standard deviation.
- Establish structures, such as data teams or professional learning communities, to support interpreting the data. Use protocols, common language, and agreements to promote relational trust and collaboration.
- Keep students at the center of the conversation. Teachers and administrators will benefit from reminding themselves that the data is telling a story about a student or group of students.
- Share ownership of the data with all stakeholders, including parents and students. By developing a partnership with families, data will become less intimidating and open opportunities to shape learning.

When you make data actionable, you make assessment matter. To fully benefit from assessment, students and teachers need to use the data to invoke meaningful change. This keeps the focus on where it should be—on student learning.

Supporting educators in their work to improve student achievement and growth is the mission that drives Kristin Moran's work at NWEA. Over 17 years of experience in the field of education—as a teacher, administrator, curriculum manager, and professional development manager—has given her a keen sense of how to support teachers in their vocation.

Assessment Just Right

Create a balanced approach to assessment to support every student

April Roethel, Senior Analyst, Assessment and Education, NWEA

What do today's educators and Goldilocks have in common? Both face big decisions to be made among a spectrum of choices. Goldilocks faced just a few decisions, and none of them impacted a classroom full of students! But just like Goldilocks, educators—from teachers to principals and district administrators—are looking for the right fit, the perfect balancing point between too little and too much. For educators, assembling assessment systems that are "just right" is a daunting challenge, but not an insurmountable one. The key lies in the concept of balance: balancing needs against resources and finding the right balance of assessment instruments and procedures. The "just right" amount of assessment takes as little time away from classroom instruction as possible, yet yields trustworthy information that supports the learning of every child. There's no one-size-fits-all approach; each school's unique needs inform what's right in each classroom.

The first step toward creating a balanced assessment plan is clearly articulating your assessment needs—this includes understanding what standards you must meet, what measures of growth matter, and what activities have real instructional impact. School administrators and support personnel, including health and counseling services, should be at the table with teachers to discuss all the issues. It's critical to engage with the whole child when considering assessment needs. Academic assessment is a crucial piece of the puzzle, but it is only one part in the whole.

The second step is understanding the full array of assessments that each class or grade currently engages with, including classroom-based assessments such as quizzes, end of course tests, research papers, performance tasks, etc. Again, all staff must collaborate to share and discuss the various assessments at play within each building.

These two steps—knowing what you need and understanding what you already have—allow you to make strategic decisions as you balance your needs with your resources. This assessment evaluation requires some level of assessment and data literacy—not only about assessment in general, but about the specific instruments that are used locally. Ask questions like:

- What kind of data do our assessments provide, and in what format?
- What instructional or educational questions do the data help illuminate?
- Who has access to the data?

When developing a balanced assessment system, the needs of every child are equally important. That's one of the reasons MAP assessments are so valuable in a balanced assessment approach. Due to its computer adaptive nature, MAP efficiently measures every student in the class, regardless of whether the student is performing on, above, or below grade level. MAP plays well into a balanced assessment system because it enables administrators and educators to:

- Identify the instructional needs of individual students
- Measure both individual and collective student growth
- Evaluate the effectiveness of teaching practices, programs, and initiatives
- Project whether a student, class, or school is on track to achieve established proficiency benchmarks

- Evaluate the impact of school programs and district-level initiatives

Students benefit tremendously in the long run when each school takes time to understand the various assessment needs, the instruments available, and how to use them. Although this sort of intentional, collaborative work can be time-intensive up front, such careful and strategic planning will help you maximize data so that students are supported as they progress on their unique learning paths. There is no such thing as an assessment system that fits everyone, but each school can create one that is just right for their students. Such a balanced approach to assessment must include multiple measures, feedback opportunities, and measurement credibility—all this will ensure your assessment approach truly supports student learning, and that's what matters most.

Having taught in the classroom for 15 years, April Roethel knows first-hand how important it is to help all kids learn. She's spent 30 years in the education field developing resources to support that mission. As both a Senior Analyst and a Senior Content Specialist at NWEA, April has relied on her experience and passion to help develop assessments that truly support teachers and students in the classroom.

Computer adaptive K – 12 MAP interim assessments help you answer a crucial question: Are my students learning?

We look forward to discussing how MAP data can help you maximize every student's learning.

Northwest Evaluation Association (NWEA) has nearly 40 years of experience helping educators accelerate student learning through computer-based assessment suites, professional development offerings, and research services. Visit NWEA.org to find out how NWEA can partner with you to help all kids learn.

References:

1. Source: Kids Count Data Center, Children in Poverty
2. Source: Paulson, Amanda, "Record Number of Homeless Children," Sept. 23,2014, Christian Science Monitor web
3. Source: American Psychological Association, 2014. "Effects of Poverty, Hunger, and Homelessness on Children and Youth"
4. Source: U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP).
5. This figure can be found at http://fcat.fldoe.org/pdf/fc06tech.pdf on page 54.
6. Reckase, M.D. "Designing item pools to optimize the functioning of a computerized adaptive test." Psychological Test and Assessment Modeling. Volume 52, 2010 (2), 127-141.
7. Gu, L. & Reckase, M.D. (2007). "Designing optimal item pools for computerized adaptive tests with Sympson-Hetter exposure control." In D.J. Weiss (Ed.), Proceedings of the 2007 GMAC Conference on Computerized Adaptive Testing, Retrieved 10/14/14 from www.psych.umn.edu/psylabs/CATCentral/
8. Carol Ann Tomlinson and Tonya R. Moon, Assessment and Student Success in a Differentiated Classroom, ASCD, 2013
9. Jan Chappuis, Rick Stiggins, Steve Chappuis, Judith A. Arter, Classroom Assessment for Student Learning: Doing It Right—Using It Well, Pearson Assessment Training Institute, 2012.

January 2015