

Balanced Assessment: Combining Multiple Measures to Drive Learning

Interim Assessment in Context

The third article in
a multi-part series
about the importance
of interim assessment.

APRIL 2014

Curriculum, instruction, and assessment form the interlocking components of educational practice. While the emphasis on one area or another shifts with time, and new methodologies transform how each gets expressed, this trio of elements serve as the foundational toolkit teachers use to spark student learning.

Once educators know what they want students to learn (curriculum), and how they want to teach it to them (instruction), assessment enters the picture. At its most basic level, assessment helps teachers and administrators understand how well students are *learning* specific material—and at the end of the process, how well they *have learned* the material.

However, in the last few decades, policy shifts have raised new questions and turned to assessment instruments to answer them. *How well are students from different socio-economic brackets learning? What comparisons can we make between different racial and ethnic groups in academic achievement?* While these are important questions to answer for a democratic society based on equality, they have imposed the use of metrics that are not strictly tied to helping individual students learn.

So what's a school district to do when faced with mixed mandates? To keep the focus on learning, many create balanced assessment systems that use multiple measures to address broad questions.

WHAT MAKES A BALANCED ASSESSMENT SYSTEM?

This series' first article, *Why Interim Assessment Matters: Interim Assessment in Context*, explored different assessment types and the purposes most often associated with them in detail. To summarize:

- formative assessment practice provides the day by day, week by week insight teachers need to make adjustments in instruction based on where students are at a given moment
- interim assessment happens at select intervals in the school year; its data provide an objective

measure of progress toward learning objectives, measure student academic growth, and inform instruction

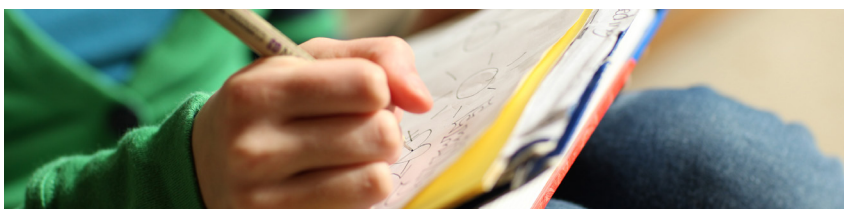
- summative assessments, whether given at the end of a unit, a semester, or—as in the case of state accountability tests, at the end of the year—demonstrate whether or not students have learned

Because formative practice and interim assessments inform learning as it's happening and summative assessments capture what's been learned, a balanced assessment system consists of all three types. When combined with additional inputs, they provide the multiple measures needed to understand student learning—and at the aggregated level, they allow educators to see achievement trends for groups of students.

What's a school district to do when faced with mixed mandates? To keep the focus on learning, many create balanced assessment systems that use multiple measures to address broad questions.

WHAT IS MEANT BY MULTIPLE MEASURES?

Multiple measures span all the instruments and sources used to gather information for the purpose of making an informed educational decision. The decision might be low-stakes, such as whether more instructional time is needed on the current topic. Other decisions might be higher-stakes, such as inclusion in special programs or courses. These sources cover a wide spectrum: they may be teacher-made and classroom-specific, or they may be psychometrically validated assessments with explicit implementation and data-collection protocols. Multiple measures should certainly be used when making high-stakes decisions that impact someone's life, such as college admissions or teacher tenure.



Here's a sampling of possible measures that can be used to make informed decisions:

- teacher observation
- principal observation
- formative strategies (i.e., exit cards)
- homework
- essays
- research papers
- student projects
- course grades
- portfolios
- performances
- teacher-made tests
- end-of-unit tests
- interim assessments
- skills diagnostics
- universal screeners
- self-assessments
- peer assessments
- progress monitoring tools
- state accountability tests
- aptitude tests
- behavioral measures (such as attendance)
- grade point average
- class rank

No single system uses all of the measures available at all times. In addition to being overwhelming to implement, it wouldn't effectively answer different educational questions such as *"Are all second graders on track to be grade-level readers by the end of third grade?"* and *"Which second graders understand phonemes?"* By using the list above to create customized sets of measures (inputs), however, teachers and administrators can get the answers they seek.

HOW A BALANCED ASSESSMENT SYSTEM HELPS ANSWER EDUCATIONAL QUESTIONS

To illustrate the usefulness of a balanced assessment system, let's examine a current policy mandate: ensuring that students are competent readers by the end of third grade.

Goal: Comprehend where current third graders stand in relation to reading competency.

Stage 1 (Administrators): Gather information

- Look at state accountability assessment scores for trends (*scores not likely to provide much additional information*).
- Look at state test data to show how last year's third graders did.
- Look at state test data for last year's second graders and make inferences about future grade 3 performance.
- Plan to give current third graders an interim assessment in fall, winter, and spring.
 - Gives a true sense of how students are performing; demonstrates the current achievement levels.
 - Permits administrators to work with principals and teachers to make program and instructional adjustments.
- Review how K – 2 students are performing in order to make any necessary adjustments to curriculum and instructional methods.

Stage 2 (Teachers): Use data to inform instruction

K – 2 teachers can cross-check their impressions of student reading readiness against an objective measure: quality data. Combining information from any skills diagnostics with their current students' interim assessment data will provide a more granular view of what each individual student is ready to learn—as well as areas where a student needs more support. They can use the data to:

- differentiate instruction
- create flexible grouping
- collaborate to implement groupings across grade in order to address common areas of need

TOWARD A MORE HOLISTIC VIEW: TRIANGULATING DATA FROM MULTIPLE MEASURES

While there are many good reasons to use multiple measures, there's also a fundamental reason to do so: all measures have some degree of error or bias. It might be acceptable to measure something like height or weight using one instrument, say a yardstick or bathroom scale, because these measurements rarely lead to any real consequences.

Generally, if the measurement isn't exact it will be *close enough*. It provides a reasonable estimate of the individual's height or weight. When making inferences about a construct as complex as the acquisition of knowledge, however, there's no one instrument that will give a "close enough" picture.

Every kind of psychometrically validated assessment has Standard Error of Measurement (SEM) associated with it. Multiple measures *that are distinct* can mitigate the bias introduced by SEM. The distinction between the measures is important, or the measurement error is simply replicated. Having multiple measures does not guarantee more measurement precision, but if the measures are intentionally chosen with awareness of the error involved in each one, greater precision is far more likely.

In considering the use of multiple measures, it's important to consider the principle of triangulation. Triangulation is simply the process of using at least three points of data when making educational decisions. Any single interim assessment score is subject to environmental or motivational influences which can affect its accuracy.

School personnel have a wealth of data available about their students. A teacher might look at a score from an interim assessment, see something puzzling, and decide to triangulate it with data from the attendance system and from a diagnostic test. This will help confirm an inference about the student—*"Ah, a string of absences when we covered this likely caused the problem, as the diagnostic data doesn't identify any issues."* The teacher might then decide that re-teaching would be a viable strategy to close the gap.

Assessment scores and the associated reports and resources provided can be extremely useful tools for informing instruction and making administrative decisions, but they should not be used in exclusion to other data sources. Using triangulation assures that the most informed and appropriate decisions are being made on behalf of each student.

Triangulation is simply the process of using at least three points of data when making educational decisions. This assures that the most informed and appropriate decisions are being made on behalf of each student.

MULTIPLE MEASURES IN THE EDUCATIONAL ECOSYSTEM

Many schools now feed data from a variety of sources into larger learning management systems (LMS), sometimes called Instructional Information Systems (IIS) or Student Information Systems (SIS). These systems are integrative: they house formal assessment data along with grades; behavioral indicators, such as attendance; samples of student work; possibly seat-time; and various other data the school deems relevant to capture. Some systems access various assessment data to generate individual student learning profiles that point the student to relevant instructional resources. These systems are in place to help teachers and administrators make sound decisions based on good information readily at hand (McCall, 2010).



The table below illustrates the different data uses at different levels within the educational ecosystem.

Table 1. Data uses at different levels within an educational ecosystem.

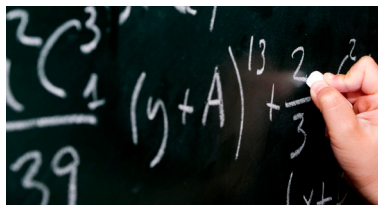
DATA USES	STUDENT LEVEL	TEACHER LEVEL	SCHOOL LEVEL	DISTRICT LEVEL	STATE LEVEL
INSTRUCTIONAL	diagnostic; growth; achievement; instructional strategies; pacing; class grouping		classroom resource allocation; trans-grade curricular alignment to learning objectives from state standards	curricular decisions and textbook adoption program evaluation	
PREDICTIVE	predicted proficiency on state summative exam; proactive intervention planning		on track for Annual Yearly Progress		
EVALUATIVE	achievement placement; course credit; grades; promotion; graduation	teacher effectiveness; performance; professional development; certification; contract renewal; tenure	school improvement planning; Annual Yearly Progress; program evaluation; principal evaluation; sanctions and rewards	Human Resource and social service allocation; professional development planning	accountability

BRINGING ALL THE MEASURES TOGETHER: COMPREHENSIVE ASSESSMENT SYSTEMS

The assessment process generating much of this data exists to improve learning for students and monitor progress toward equitable outcomes for different groups of students. Its function is to shed light on programs, policies, and practices with the intent to improve them. Information gained from assessment can protect the vulnerable; promote innovation; and address social inequities. Its power for positive impact is enormous. “Although assessment is often seen as a tool to measure the progress of individual students, it also allows individuals, communities, and countries to track the quality of schools and educational systems.” (Braun, 2006)

Of course, fulfilling this role requires a balanced approach to assessment, and finding this balance begins with developing a coherent assessment plan. Like nesting Russian dolls, assessment plans exist within assessment plans. Teachers will develop assessment plans for their classes to ensure that the instructional process is working. These classroom-based plans exist within the school’s assessment plan, which exists within the district’s assessment plan, which exists within the state’s assessment system.

School systems administer multiple assessments to answer different questions and meet state and federal mandates. These systems should be as efficient as possible, and here again, interim assessment can play a crucial role. A well-constructed interim assessment, with a stable scale and deep item pool that can adapt for students who are on, above, and below grade level, can fulfill multiple purposes within a comprehensive system.



Developing a coherent and balanced assessment system calls for vision, leadership, collaboration, communication, and open-mindedness to change.

- For students, interim data can help them work with teachers to set learning goals and monitor progress toward their completion.
- For teachers, interim data can provide an objective measure of student achievement, demonstrate growth, point to instructional opportunities, and inform differentiation strategies.
- For administrators, interim data can show trends for grade levels, schools, and classrooms. Data can be used to evaluate the efficacy of programs and inform evaluation rubrics, inform projections of performance on state accountability tests, and demonstrate the effectiveness at efforts toward closing achievement gaps.

Undeniably, developing a comprehensive assessment system is hard work. Fortunately, there are tools

and procedures expressly designed to facilitate this work. But tools don't replace sound professional judgment: they support it. Developing a coherent and balanced assessment system calls for vision, leadership, collaboration, communication, and open-mindedness to change.

Internal versus external reporting requirements, internal needs, and formative, interim, and summative purposes help frame the effort. **Focusing unwaveringly on students and what will best help them keeps the work grounded in what matters most.**

There is no such thing as a perfect assessment system, but one founded on a theory of action that incorporates multiple measures with the clearest signals, provides opportunities for relevant feedback, and has credibility and defensibility (Gong, 2011) will go a long way toward maximizing data so that students are supported as they progress on their unique learning paths.

This article is the third in a multi-part series. In the next part, we'll cover the top 10 questions to ask when comparing and contrasting interim assessments.



Founded by educators nearly 40 years ago, Northwest Evaluation Association™ (NWEA™) is a global not-for-profit educational services organization known for our flagship interim assessment, Measures of Academic Progress® (MAP®). More than 6,800 partners in U.S. school districts, education agencies, and international schools trust us to offer pre-kindergarten through grade 12 assessments that accurately measure student growth and learning needs, professional development that fosters educators' ability to accelerate student learning, and research that supports assessment validity and data interpretation. To better inform instruction and maximize every learner's academic growth, educators currently use NWEA assessments and items with nearly 10 million students.

References

- Braun, Henry; Kanjee, Anil; Bettinger, Eric; Kremer, Michael.** *Improving Education through Assessment, Innovation, and Evaluation*. American Academy of Arts and Sciences; 2006. Retrieved on February 28, 2014: <https://www.amacad.org/publications/braun.pdf>
- DePascale, Charles A.** *Multiple Measures, Multiple Meanings*. Presented at the Reidy Interactive Lecture Series; Center for Assessment; Boston, MA; September 2011. Retrieved on February 28, 2014: http://www.nciea.org/publication_PDFs/RILS_CD2011.pdf
- Goertz, Margaret E.** *Multiple Measures, Multiple Uses*. Presented at the Reidy Interactive Lecture Series; Center for Assessment; Boston, MA; September, 2011. Retrieved on February 28, 2014: http://www.nciea.org/publication_PDFs/RILS_PG2011.pdf
- Gong, Brian.** *Reidy Interactive Lecture Series' Multiple Measures: a personal response*. Presented at the Reidy Interactive Lecture Series; Boston, MA; September, 2011. Retrieved on February 28, 2014: http://www.nciea.org/publication_PDFs/RILS_BG2011.pdf
- Lindquist, E.R., Editor.** *Educational Measurement*. American Council on Education; Washington, D.C.; 1951. Retrieved on February 28, 2014: http://archive.org/stream/educationalmeasu00lind/educationalmeasu00lind_djvu.txt
- McCall, Marty.** *Interim Assessments: Why now? What next?* Presented at the Annual Meeting of the National Council on Measurement in Education; Denver, CO; May, 2010.