



Ronald S. Thomas

*Terms every board member needs to know
to understand complex data*

101

Testing and Assessment

Do data and assessment terms confuse you? Can you tell the difference between a norm- and a criterion-referenced test? What are benchmark data and why are they so important? What are school abilities tests? What is Adequate Yearly Progress (AYP)?

As your curriculum team clicks through these various phrases, are your eyes starting to roll up in the back of your head?

Don't worry; you are not alone. Understanding assessment data is a complex task, but it's one that we can make easier. While this article can't answer all your questions about assessments, it will give you some of the background you need to be an informed board member.

Nature of assessment data

Assessment data are observations or facts that must be collected, organized, and analyzed to become useful. The key point to remember is that data are merely numbers or words that do not have meaning in and of themselves.

Data only acquire meaning when educators and board members sift the numbers through their personal experiences and place them within the context of their school or the district as a whole. Because your personal and social filters are different, don't be surprised if your perception varies from the superintendent and school staffs.

At one time, schools lacked sufficient assessment data to make good instructional decisions, but now many are snowed under with information. Organize your data discussions around a series of questions, such as:

- Where do assessment data come from?
- Why do schools collect assessment data?
- What, exactly, is being assessed?

■ Against what criteria are assessment results being measured?

Asking each of these questions provides a useful framework to increase your understanding of what you are discussing.

Where do assessment data come from?

According to Jonathan Supovitz and Valerie Klein, authors of the widely respected *Mapping a Course for Improved Student Learning*, there are three major sources of assessment data: external data, school-wide or district-wide benchmark data, and classroom data.

External assessment data are standardized norm- or criterion-referenced tests that originate and are scored outside the school. Terra Nova, SAT, Stanford 10, and each state's No Child Left Behind (NCLB) test are examples. Results from external assessments can suggest an initial focus for a school's attention, but they are not designed to be frequent enough or to provide specific enough data to give precise instructional guidance for a whole year.

School-wide or district-wide benchmark data are collected frequently and systematically across an entire grade, content area, or course, perhaps over an entire school district, at several predetermined points in the year. These common assessments can provide guidance for instructional adjustments, interventions, and professional development.

Most importantly, if scored and analyzed collaboratively by teachers, discussion about common assessment results can reinforce a culture of inquiry, based on data, among the staff. However, these are often the most underutilized type of data in a school because teachers don't have the time to analyze them or the training and experience to collaborate.

Classroom assessment data include quizzes, unit tests, essays, performance assessments, and personal communica-

tions developed and administered by individual teachers. Supovitz and Klein liken classroom assessment data to having a personal global positioning system (GPS) for each student. Classroom assessments provide the opportunity to provide quick and flexible feedback to each student throughout the year and to make immediate adjustments in instruction.

Each source serves a different purpose, and reports on student achievement from the three often vary significantly in

how they portray data. Successful schools use an established system to bring together the results of all three types when making instructional decisions.

Why do schools collect assessment data?

There are two important uses for the assessment data that schools collect. One is for educational accountability (or, in other words, to prove to the public that education is doing its

New assessment tools help analyze learning, refine teaching

Joetta Sack-Min

Forget about filling in circles with No. 2 pencils and waiting months for test scores. Dozens of new technologies are making their way into classrooms to help analyze student learning.

These tools, most of which provide instant feedback, seek to give teachers more time to spend on instruction rather than grading and evaluating student work. Used correctly, some assessment devices can give teachers a more detailed look at how students are progressing, what areas require more time, and where they should adjust lessons or teaching strategies.

"The real potential is to be able to meet a student and meet their needs," says Mary Ann Wolf, executive director of the State Educational Technology Directors Association. "We are seeing growth and jumps in these different technologies. These are more comprehensive systems, more formative, and seem to be improving and connecting to the standards."

But assessment technology won't take the place of a savvy teacher, experts agree. Some caution that these systems may require more work up front by administrators and teachers to understand the systems and the data they produce.

"You can't assume that because you have the databank that teachers can interpret the data correctly or modify the curriculum without adequate training," says Don Knezek, the chief executive officer of the International Society for Technology in Education. "This may require a culture change on campus, and that's where leadership comes in."

The following is a partial list of some

types of devices to use in assessing student performance.

- A "student response system," such as Quizdom, allows students to record their quiz or test answers on a handheld device that instantly grades and delivers an individual score or class report. The advantages are instant scores, less grading time for teachers, and a format that ensures that everyone participates. Students in the back of the class can't slack off, and shy students have a voice. The systems provide class reports so teachers can see the questions or areas where students struggled.

- Writing samples and essays can take days to grade, so many schools are now turning to computer-based programs. Students type an essay or assignment, and the software grades the punctuation, grammar, language use, and organization of the sample. One popular program is Vantage Learning's My Access!, which compares a student's sample to hundreds of others and gives a written review and grade of its focus, content, organization, language, and mechanics.

- More recent models of computerized whiteboards, which are becoming commonplace in classrooms, include interactive functions that can be used for informal assessments, such as polling students.

- Videos and interactive software are designed to better engage students and can also assess their critical thinking and problem-solving skills. ThinkLink, a subsidiary of

Discovery Education, has designed a series of episodes with characters that need help with real-life problems. Students must extrapolate data and perform mathematical equations to find solutions.

- Other programs present data and materials for students to analyze as part of a more comprehensive test. The Educational Testing Service offers the *iskills* Assessment, which requires students to glean information from a database, develop a spreadsheet, or compose presentations explaining their findings as part of an online assessment. Learning.com offers a similar product, the TechLiteracy Assessment, which also delves into social and ethical issues related to the use of technology.

While it may seem that artificial intelligence has taken over assessments and grading, these devices may only be the beginning (although most experts agree that bubble tests will be around for a few more years, so don't throw out the No. 2 pencils yet).

Knezek says more assessment will be based on the technologies that schools often deem to be nuisances—cell phones, PDAs, and other devices that allow text messaging and access to social networking sites such as Facebook.

"Some of the things people are really fearful of—we're just going to have to figure out how to use those effectively and safely, or we'll be putting students who don't have access to technology at a disadvantage," he says.

Joetta Sack-Min (jsack@nsba.org) is associate editor of *American School Board Journal*.

job). State and national external assessments are considered the most important for accountability purposes, while classroom assessments are the least important.

The second major use of assessment data is for instructional decision making—data collected explicitly to improve the job that educators are doing. State and national assessments are the least helpful here. Classroom assessments provide the most useful ongoing data that teachers need to impact instruction, while common assessment data help teachers benchmark student progress against peers in other classes or schools.

To have an effect on the NCLB state test (accountability data), benchmark and classroom assessments (instructional improvement data) must measure student performance using the same content, format, and level of rigor. This will not happen automatically. Teachers need carefully designed assistance to align their classroom assessments with the state standards.

What exactly is being assessed?

There are four general types of external assessments.

General achievement tests assess the wide spectrum of student knowledge and skills. These tests usually take several hours on multiple days to administer. However, the tests can provide data on vocabulary word analysis, listening and reading/reading comprehension, language (such as spelling, capitalization, punctuation usage, and written expression), math (concepts estimation, problem solving, and computation), science, and social studies. Each commercial series—such as the Iowa Tests of Basic Skills, Terra Nova, and the Stanford 10—includes aligned tests for all grade levels.

Special area achievement tests center on one subject area, such as reading or mathematics. Because of this focus, they can provide more detailed information, such as data on phonemic awareness, decoding skills, letter/sound correspondence, word meanings, and fiction and nonfiction reading comprehension. The Gates-MacGinitie Reading Test is a well-known example. Most state NCLB assessments fall into this category because they focus solely on reading and mathematics.

A subcategory of special area achievement tests are diagnostic assessments. While all assessments have some diagnostic value, these tests are specifically designed to highlight the strengths and weaknesses of students in greater detail. The Stanford Diagnostic Reading Test, the Diagnostic Assessment of Reading, and computerized tests such as the STAR Reading and Math Assessments are examples.

The first three types of assessments are all achievement tests because they provide data on the knowledge and skills students can demonstrate. School abilities tests, which measure abstract thinking and reasoning abilities, provide information on students' potential to learn in school.

Tests such as the Cognitive Abilities Test (CogAT) and the Otis-Lennon School Ability Test (OLSAT 8) measure verbal abilities, such as sentence arrangement, verbal classification, and verbal analogies, as well as nonverbal abilities, including figure

classification, pattern matrices, and number inference. Results may be correlated with achievement test data to relate students' actual achievement to their ability to perform in school.

Against what criteria...?

Assessments are standards- or criterion-referenced (CRTs), norm-referenced (NRTs), or a combination of both.

On CRTs, student performance is measured against the content standards (the criteria) that all students should know and be able to do. Standards have been written by each state, usually modeled after national standards documents developed in the 1990s. Some state content standards are very precise and concrete, while others are more global and vague. All test results are reported as the percentage of students at the advanced, proficient, and basic levels in relation to the state's content standards.

States also vary in their performance standards, or the percentage of students the state expects to be in the advanced and proficient levels at each school each year up to 2014. Also called the "cut score," this percentage is known as the annual measurable objective (AMO) in reading and math.

Each state has different content standards, a different state test, and a different "cut score." It is, therefore, impossible to compare scores from state to state on their NCLB tests.

AYP is based on the combined school performance for all grades on the state's NCLB tests. AYP is calculated separately for reading and math. A minimum percentage of students must score "proficient" or "advanced" (at or above the AMO) each year for the school to meet AYP.

The percentage needed to meet AYP will increase each year until it gets to 100 percent. In 2014, under the present NCLB law, all students are expected to meet content standards by scoring at the advanced or proficient levels. All subgroups in the school (including all races, special education students, and children living in poverty) must score at or above the annual measurable objective for the school to meet AYP.

Norm-referenced tests (NRTs) compare student performance to that of a similar group of students who took the test when it was normed—the "norm group." This may have been several years before. Results are often expressed as percentiles (the percentage in the norm group that scored lower than students in your school). Many NRT test developers also report standard-based data in relation to major objectives included in the test.

Too many concepts to make much sense? Many school leaders are struggling with the same issues. As we work together to sort out the uses, value, and importance of assessment data, we can be in a better position to use them effectively and capture the power of data-based decision making. ■

Ronald S. Thomas (rathomas@towson.edu) is the associate director of the Center for Leadership in Education at Towson University in Towson, Md. He has worked with more than 100 school improvement teams over the past 15 years.