Part 1: Primary Tools

Testing, or "assessment," plays a vital role in education today. Test results are often a major force in shaping public perception about the quality of our schools. As a primary tool of educators and policy makers, assessment is used for a multitude of purposes. Educators use assessment results to help improve teaching and learning and to evaluate programs and schools. Assessment is also used to generate the data on which policy decisions are made. Because of its important role, educational assessment is a foundation activity in every school, every school district and every state—a vital component in innovation, higher standards and educational excellence.

Testing has been a pivotal part of American education since early in this century when educators began to seek more reliable and valid means to evaluate students and programs. In the past 40 years, there has been explosive growth and profound change in education. At every step of the way, educational assessment has responded with innovation in measurement and technical expertise. In the past ten years alone, the field of testing has undergone tremendous change because of the emphasis on education reform and development of new education standards.

Local and state education agencies are called upon today, to make many crucial decisions regarding how students and programs are assessed—decisions often involving significant time, effort and public resources. Making the right decisions about testing begins with having a basic understanding of the need for assessments that are valid, reliable and fair, and that fulfill their designed purposes. Though testing is often perceived as a technical field, these "basics" of assessment are not difficult.

This information addresses those "basics." For more than seventy years, CTB/McGraw-Hill has worked in partnership with school districts and states to create successful assessment systems. We hope this information will serve not only as a starting point but also as a continuing reference tool for local and state school board members, educators and policy leaders seeking a firm footing in assessment.

Part 2: The "Basics" of Educational Assessment—Why Test?

Every teacher and parent has heard a student ask the question, "Why do we have tests?" This is the most fundamental question in educational assessment, and it has multiple answers. Assessment is used to:

Monitor educational systems for public accountability
Help provide information to better identify instructional practices
Evaluate the effectiveness of instructional practices
Measure student achievement
Evaluate students’ mastery of skills

Given the different uses for assessment, it is critical that educators select the appropriate type of test. Before examining the various kinds of assessments and the information they provide, let’s first consider the principles that guide assessment development and use.

Four Principles to Consider

Although educational testing is a complex field, four basic principles provide a foundation for further understanding:

1. Standards First, Then Testing.

When states and communities reform their education systems, a logical sequence of events must be followed. First, the goals for each education system must be set. Second, standards must be adopted to outline what children should know and be able to achieve. These standards should be written in a way that will help students meet the stated goals. Following the adoption of standards, curricula must be set and instructional materials selected to help teachers assist their students in meeting the standards.

Finally, assessments are developed to measure student progress toward meeting the standards. In other words, assessment should follow, not lead, the movement to reform our schools. As we continue to find ways to improve education, it is important for educators and policy makers to use a sequence that starts with goal setting and ends with assessment. Only then can we build and use new tests that accurately measure our progress toward meeting standards.

2. Tests Measure Educational Progress--They Don’t Create It.

The purpose of testing is to deliver accurate and reliable information, not to drive educational reform. Some politicians and policy makers have suggested that new tests alone will create higher levels of educational achievement. What they are really looking for is better results. It is important for school administrators and policy makers to understand that a new assessment system cannot cure ailing education systems. Tests do not create better students; good teachers and good schools do! The problems facing our nation’s schools are serious. There is no single cause, and therefore no simple cure for these problems. There are no shortcuts to improving student achievement and creating a world-class workforce. We continue our search for ways to improve student achievement, not rush into thinking that a new testing system will create better schools.

3. No Single Test Does Everything--The Importance of Multiple Measures.
No single test can do it all. A diagnostic test to determine the emission level of an automobile engine will not tell you that the tires need air. A different procedure is needed to provide that information. The same goes for tests in education. No single test can ascertain whether all educational goals are being met. A variety of tests, or "multiple measures," is necessary to tell educators what students know and can do. And just as different tests provide different information, no one test can tell us all we need to know about one student's progress. This "multiple-measures approach" to assessment is the keystone to valid, reliable, fair information about student achievement. Any one type of test, whether norm-referenced, multiple-choice or performance assessment, is only one part of a balanced approach to assessment.

For example, some tests are designed to indicate whether a student needs additional work in specific subjects, while others measure overall group progress toward broadly stated goals. Because curricular emphases differ from state to state, as do the purposes of testing, a multiple-measures approach means that states and local school districts often use different types of tests to assess students. Educators understand the real power and utility of creating testing programs that combine performance assessments, norm-referenced tests and other measures. This approach puts the right kind of assessment to work for the right purpose. Performance assessments, for example, might be used for instructional purposes, while norm-referenced tests are used to generate comparative information. Such data continue to be in great demand as the educational community seeks to build greater accountability measures into their educational systems.

4. The Importance of Valid, Fair, and Reliable Assessments.

All tests and test types, whether standardized, multiple-choice, or performance assessment, should be held to the same high technical standards for producing accurate information. No test should be selected and administered without first determining how its results will be used and its appropriateness to the subject matter. Furthermore, no test should be used without reviewing its technical strengths, including fairness, validity, and reliability. All assessments should be designed, piloted, and published using nationally accepted technical standards such as those developed by the American Psychological Association, the American Educational Research Association, and the National Council on Measurement in Education. In recent years, many new assessments and test formats have been developed. These tests, too, must be held to the same high standards. Unvalidated tests, especially those with high-stakes outcomes, should not be administered.

Part 3: Kinds of Tests

Each day millions of American school students take tests. Over 95% of these exams are "pop quizzes," oral presentations, or some other type of teacher-made test. However, standardized assessments developed by test publishers—the type of test that best evaluates student learning over time in comparison with others—usually receive the most attention. Typically, such tests are both standardized and norm-referenced. They are used only once or twice a year, and provide objective information about each student's progress in mastering the school curriculum.
For many years, educators and the public perceived standardized tests as exclusively norm-referenced, multiple-choice examinations. That was not exactly true then and it certainly is not true today. A standardized test is one that is always given in a consistent manner, with the same directions, the same questions, and the same time limits. Thus, scores can be compared with confidence in test validity and reliability. All assessments administered within a state or local testing program should be standardized, no matter what type: performance based, norm-referenced, or criterion (standards) referenced.

Educators recognize the value of using a variety of tests. A comprehensive assessment program may include several different measures, among them the following basic types and formats:

**Standardized achievement tests**

These tests are commonly used to provide valid, reliable, and unbiased information about students' knowledge in various areas. "Standardized" means that the test is always given and scored the same way. The same questions are asked and the same directions are given for each test. Specific time limits are set, and each student's performance may be compared with that of all the other students taking the same test. Most standardized achievement tests are norm-referenced, multiple-choice tests.

**Norm-referenced tests**

Norm-referenced achievement tests measure basic concepts and skills commonly taught in schools throughout the country. These tests are not designed as precise measures of any given curriculum or single instructional program. Results from norm-referenced tests provide information that compares students' achievement with that of a representative national sample. This gives teachers the opportunity to compare their students with other students. So, when a teacher says that a student scored at the 82nd percentile, that student's score was equal to or better than 81 percent of the scores of all the students who took the same norm-referenced test during the norming process.

**Criterion-referenced tests**

This type of assessment is designed to compare a student's test performance with clearly defined curricular objectives, skill levels, or areas of knowledge. While norm-referenced test results compare student performance to peers—for example, a student spelled better than 95 percent of his or her classmates—results from criterion-referenced tests compare the performance to a predefined set of objectives—and demonstrated mastery (knowledge) of a specific subject, such as long division.

**Multiple-choice tests**

Many standardized tests give students the opportunity to select responses to test questions among a number of specific choices. This format, called "selected response" or "multiple choice," is efficient and practical. Carefully designed multiple-choice questions can provide valid information about students' knowledge and their ability to reason logically and apply complex thinking processes.
to solve problems. Norm-referenced tests are usually administered in a multiple-choice format, where the correct answer is provided along with incorrect answers. These are the tests most adults remember taking in their youth. In most instances, multiple-choice tests are scored by computers and provide impartial, accurate results.

Performance assessments

Performance assessments are types of tests that directly assess pupil performance. Students may be asked to write an essay or short response, draw a conclusion, respond to a reading passage, or perform a science experiment. Teachers or other school personnel observe students’ performances and rate the outcomes. This kind of assessment is also useful in measuring listening skills, writing, and the process of problem solving. Performance assessments can also be standardized so that the test is given and scored the same way at each administration.

Part 4: The New Era of Standards and Assessments

In the late 1980s and throughout this decade, setting the National Education Goals and developing standards in science, mathematics, English language arts and history spurred a number of states and local districts to re-examine their testing programs. In many cases, states and districts revised their programs or created new ones to reflect the standards. For example, creating new performance standards by some states led to the development and use of new performance assessments. At the same time, performance assessments became increasingly popular because of their ability to generate and capture rich information about the cognitive progress of students.

These changes have led to public debates among educators, reformers and legislators about the utility of various assessments. In these debates, one type of assessment is often pitted against another in an effort to determine which is the "best" test. It has been argued by some reformers and cognitive scientists that higher-order skills involve mental processes that are difficult to translate into conventional assessments such as norm-referenced tests. While this may have been true in the past, testing professionals are now able to build multiple-choice tests that measure higher-order as well as basic skills. Well-designed, standardized tests can provide reliable information and trend data on various student skills over time.

The winners in this debate are those who understand the use of multiple measures. Remember: No single test does it all. Testing in American Schools: Asking the Right Questions, a national report issued in 1992 by the Office of Technology Assessment of the U.S. Congress, sums it up best: To outsiders listening in on this debate, it may appear that proponents of conventional and new forms of assessment are adversaries locked in an intractable stalemate. Closer inspection, however, reveals that testing policy is not a zero-sum game in which either existing testing or new methods win, but an arena with multiple and mutually compatible choices.
The key is using the kind of assessment that best provides the desired information. Thus, although some activists in the debate have carved out extreme positions, most agree on at least two other fundamental points:

Different forms of testing can, if used correctly, enrich our understanding of student achievement. Tests of any kind should be used only to serve the functions for which they were designed and validated. On this common ground, the OTA report concludes, it may be possible to build genuine reform.

Part 5: Designing Local and State Assessment Systems

Designing local and state assessment systems can be a complex and arduous task. Substantial time and effort are necessary to create assessment systems that are valid, fair, reliable and that perform the desired function. Yet, like so many other aspects of assessment, there are basic steps that inform the design process and facilitate implementation of testing programs.

Step One: Set Assessment Goals and Objectives.

At the outset, educators and school board members must determine their assessment goals. Do they hope to determine student progress? Evaluate programs? Link instruction directly to assessment? The district or state must also determine the type (or types) of information wanted from an assessment system and whether a new assessment augments an existing one. Districts, in particular, must look at how a new district-wide assessment system complements a state testing program. Once educators and board members answer these questions, they can begin to consider which kinds of assessments will meet their needs. During this process, the principle of "sequencing"—establishing goals, setting standards, developing curriculum and then designing assessments—is very important. Finally, educators, administrators, and school board members must take extra care to listen to the recommendations and concerns of parents, teachers, and students.

Step Two: Secure the Help of Assessment Professionals.

Test publishers like CTB/McGraw-Hill work with school districts and states throughout the country to design and implement assessment programs. As educators and measurement professionals, they bring with them years of experience and expertise in all areas of assessment: design, technical specifications, reporting, scoring, and professional development. Assessment professionals will assist a district or state in performing a number of critical functions. At the outset, they will help clarify and refine assessment goals and objectives. Then, they will partner closely with the district and state during the next three steps.

Step Three: Design the Program.

Assessment professionals will work with district and state to turn assessment goals and objectives into an assessment plan. As part of this design, appropriate grade levels for assessment and frequency of testing must be determined. In addition, an assessment format and the types of tasks included in the new test are to be selected.
Step Four: Develop the Assessment.

Assessment professionals work with district and state to create the actual assessment tasks and field test (pilot) trial assessments. They will also develop scoring and reporting procedures. Depending on degree of complexity and customization, these activities can take several months or even several years. Along the way, there can be numerous reviews of design, assessment tasks, scoring and reporting systems.

Step Five: Implement Professional Development Activities.

While many teachers and administrators have been involved in the design process, many others have not. It is critical that teachers and principals be trained to understand and administer the test. They must also learn how to explain test results to students and parents. If a customized performance/portfolio assessment is used, special training will be needed to help teachers use the results to improve instruction. In addition, a special cadre of teachers may be needed to help score the assessment. Assessment professionals can assist with all these tasks.

Step Six: Implement the Assessment.

Following the design, development and field testing of the assessment, the new system is ready to be administered to students. In some cases, particularly in state programs, full-blown implementation is often preceded by an "interim assessment" in which the test is piloted while research and development continue. In either case—"interim assessment" or full implementation—districts and states must carefully monitor the use of the test. They should also be sure that there is solid knowledge and understanding of the assessment among educators, parents, students and community leaders. In some cases, test publishers have assisted states in creating a package to build public awareness and support for new assessments.

Step Seven: Evaluate the New System.

Together, the assessment professionals and the district or state must evaluate the new assessment system. Is it accomplishing the desired goals and objectives? In what ways should the test be refined or altered? Should it be expanded to additional grade levels? What type of continuing professional development is needed? All these questions should be answered in the evaluation.

Part 6: Summary

The role of tests in our schools requires an understanding of assessment practices and principles on the part of school board members. Policy decisions and debate that accompany the design and implementation of new assessment programs often call for the active participation of school board members.
CTB/McGraw-Hill believes that the basic principles and steps described here will help you as your local district or state develops new assessment programs. Despite the technical complexities of assessment, we believe that you need first consider only a few fundamental questions to guide the formation of a new assessment system:

What does your district or state hope to achieve with a new assessment?
What is the purpose of the new assessment system: to measure student progress, to evaluate programs, or to determine accountability?
On what education goals and standards will the assessment system be based?
Have steps been taken to ensure that the new assessment is valid, fair and reliable?
What type or types of information does the district or state hope to generate from the assessment?
How will information from the new assessment be used?

We hope this information will be helpful as you participate with your district or state in considering the selection or development of a new assessment system.

Copyright © 2013 CTB/McGraw-Hill LLC. Subject to Terms of Use. Read our Privacy Notice.