Introduction

> For assessments to become an integral part of the instructional process, teachers need to change their approach in three important ways: They must (1) use assessments as sources of information for both students and teachers, (2) follow assessments with high-quality corrective instruction, and (3) give students second chances to demonstrate success. What makes these changes in approach so difficult, however, is that each change compels teachers to depart significantly from the practices they experienced as students. In other words, teachers must think about and use assessments differently than their teachers did. (Guskey, 2007, pp. 16–17)

For years, the area of assessment was relegated to a secondary role in the educational process. Many educators feel assessment was ignored, misused, and totally misunderstood by administrators, teachers, parents, and students. In the last decade, however, assessment has emerged as one of the major components in the restructured school. One cannot open an educational journal, attend a district workshop, or listen to media reports without hearing about standards-based reform, high-stakes standardized tests, and international testing results.

The emergence of authentic assessment coincides with the ever-increased emphasis on standardized testing. Almost everyone is aware of the controversy surrounding standardized tests. Charges that high-stakes standardized tests do not always measure significant learner achievement, do not focus on thinking skills, and do not accurately reflect students' understanding of important concepts have increased as the variety and number of required tests have increased.

STANDARDIZED TESTS AND CLASSROOM ASSESSMENTS

Standardized Tests

Despite criticisms that standardized tests do not always assess what students are learning, and that their emphasis is on mostly factual knowledge rather than higher-order thinking and application, they are still the yardstick that the public and policymakers use to measure educational progress. Standardized tests are viewed by many people as being valid and reliable and, for the most part, the most effective method to compare students, schools, districts, states, and countries.

Standardized test scores are used to determine many important educational decisions. States use high-stakes standardized tests to promote or retain students, award diplomas, reward administrators and teachers with bonuses if their students perform well, or to fire teachers and school administrators and close schools if students perform poorly. Standardized tests are required for admission at most colleges and graduate schools, and they are used to help determine certification in areas such as law, education, medicine, and accounting.

Recently, however, improvements in terms of tested achievement in the K–12 educational system have reached a plateau. According to Hargreaves and Shirley (2008), "The curriculum is shrinking, classroom creativity is disappearing, and dropout rates are frozen. Top-down prescriptions without support and encouragement at the grass roots and local level are exhausted" (p. 136). Hargreaves and Shirley believe that the data on the existing strategies and the economic need for increased innovation and creativity have necessitated a shift in education reform. They warn that "High-stakes and high-pressure standardization, where short-term gains in measurable results have been demanded at any price, have turned many U.S. schools not into learning-enriched environments, but into enervating 'Enrons' of educational change" (p. 136). With the collapse of the financial sectors of the world, the need for economic innovation and creativity has never been greater and the shift in educational reform will, hopefully, reflect the need for significant educational changes to meet the increasing challenges of the twenty-first century.

Classroom Assessments

Although research-based instructional strategies and classroom management strategies are critical components of teaching, the research on classroom assessment indicates that it is one of the major factors that improve student achievement. Marzano (2006) cites the research by Paul Black and Dylan Wiliam (1998) synthesizing more than 250 studies conclusively showing that formative assessment does improve learning; moreover, the gains are among the largest ever reported for educational interventions. Marzano says, "To the surprise of some educators, major reviews of the research on the effects of classroom assessment indicate that it might be one of the most powerful weapons in a teacher's arsenal" (p. 2).

Teachers usually develop most classroom assessments. These formative assessments consist of a variety of methods including logs, journals, debates, graphic organizers, projects, products, performances, experiments, portfolios, critical or creative writing assignments, skill tests, etc. The purposes of these assessments are to provide feedback for teachers and students, evaluate students' knowledge and understanding of key concepts and standards, and guide the instructional process by differentiating teaching to meet the diverse needs of all learners. Formative assessment provides information to both the teacher and the student about student progress toward learning goals, and the constructive feedback encourages students to self-assess and adjust in order to improve. Effective classroom assessments that are frequent and integrated seamlessly with instruction provide a continuous feedback loop that *informs* instruction.

ASSESSMENT LITERACY

Stiggins (1994) discusses the need to develop "assessment literacy" among all the stakeholders concerned about the quality of schools and the achievement of students. He describes assessment literates as those who understand the basic principles of sound assessment and how this relates to quality instruction. Along with this, teachers must strive to maintain a balanced use of assessment alternatives.

Stiggins says that the educational system will continue to use both standardized testing *and* classroom assessment. Assessments will continue to provide valuable information for important decision making, but they are also valuable teaching tools that should be used to promote meaningful learning for all students.

ACCOUNTABILITY

Elmore (as cited in Crow, 2008) believes that accountability in education today refers to systems that hold students, schools, or districts responsible for academic performance. He believes the current accountability system has further devalued the professional knowledge of the field. Elmore says the consequences of relying so heavily on testing and sanctions reinforces the idea that, "educators already know all that we need to know to solve these problems, and the reasons why schools aren't performing is that educators are just contrary and incompetent people" (p. 46). He recommends that people running the accountability system in the No Child Left Behind (NCLB) legislation review accountability models in Canada, Australia, and Europe in which schools are given feedback on the performance of their students, and they are also given support and challenges to improve. He believes the United States is incredibly overinvested in testing and sanctions, and incredibly underinvested in capacity building.

People know when students and schools are doing poorly, but they need to know the process from taking a school from point A to point B and what organizational structures and resources are necessary. Elmore (as cited in Crow, 2008) states:

> It's not the policy makers who are going to make this period of educational reform successful—it's the people on the ground who are going to do it. They don't have all their best ideas stashed away in some desk drawer somewhere—they're doing what they know how to do. If they're not doing the right thing, we need to figure out how to put them in a situation where they can learn how to do it differently. (p. 47)

The importance of providing high-quality professional development to teachers has never been greater. Beginning teachers, veteran teachers, and second-career teachers all need to experience hands-on training and coaching on the best practices of staff development so they will be able to take schools from point A to point B, regardless of the socioeconomic and academic challenges facing the students.

GRADING

Grades are, unfortunately, an integral part of the American educational system. As early as kindergarten, students receive grades that they might not understand. Ask teachers what they hate most about teaching, and there's a good chance the answer is "giving grades." Many a teacher has agonized over report cards, trying to decide the fate of a student. It is a gut-wrenching task for teachers to translate everything they know about what a student knows, can do, and feels into one single letter or numerical score. That final grade may determine promotion or retention. It may determine placement in a class, school, or participation in an extracurricular activity. It may determine honor roll, class ranking, college admission, college scholarship, or career placement. Currently, in some states such as Georgia and Louisiana, a student's average could prevent him or her from obtaining a scholarship to four years at a state university. Grades are, indeed, high stakes for students and their families. Many important decisions are made on the basis of a grading system that can be inconsistent, arbitrary, and, sometimes, punitive.

Grades affect the self-confidence, self-esteem, motivation, and future of a student. Fortunately, some school systems are moving away from traditional letter and number grades at some levels and adopting performance indicators developed from the standards on report cards. They are also using portfolios, student-led parent-teacher conferences, anecdotal records, checklists, multiple scores, and other more authentic descriptors of a student's progress. But despite attempts to restructure report cards to reflect the emphasis on performance, standards, thinking skills, and other indicators, traditional As, Bs, Cs, Ds, and Fs are still used to pass judgment on students.

With the stroke of a pen or the "bubble" of a Scantron sheet, a teacher can pass judgment on a student. "It [a grade] marks the lives of those who receive it. It may not be imprinted on the forehead, but it certainly leaves an impression" (Majesky, 1993, p. 88). The grade can become the scarlet letter of Puritan days—especially if it is based on trivial tasks or inappropriate behavior, absences, attitude, and punctuality. "As at the last judgment, students are sorted into the wheat and the chaff. Rewards of As and Bs go out to the good, and punishments of Fs are doled out to the bad. 'Gifts' of Ds (Ds are always gifts) are meted out, and Cs (that wonderfully tepid grade) are bestowed on those whose names teachers can rarely remember" (p. 88). Another challenge facing teachers is how their classroom grades correlate to the state's standardized test scores. Parents tend to be displeased if Mary has earned all As in Mrs. Brown's language arts class, but is retained and must attend summer school when she fails the state test. The *disconnect* between "teacher" grades and "standardized test" grades is difficult to explain to the public.

Grading is indeed a complicated issue, and used for accountability, an issue that must be addressed. Airasian (1994) says that grading means "making a judgment about the quality of a pupil's performance, whether it is a performance on a single assessment or performance across many assessments" (p. 281). The judgment, however, should be based upon clear criteria, correlated to course outcomes and state standards, in order to be valid. Arbitrary

judgments based on personality, behavior, and effort distort the grade and the inferences that are made from the grade.

O'Connor (2002) believes communicating student achievement is the primary purpose of grades. "Simply stated, if clear communication does not occur, then none of the other purposes of grades can be effectively carried out" (p. 16). Students, parents, and administrators must all be aware of the purposes of the grades and the criteria upon which they are based.

TRADITIONAL COGNITIVE SCIENCE

The methods of assessment used in schools are often determined by beliefs about learning. Early theories of learning indicated that educators needed to use a "building-blocks-of-knowledge" approach whereby students acquired complex higher-order skills by breaking learning down into a series of skills. Every skill had a prerequisite skill, and it was assumed that after the basic skills were learned, they could be assembled into more complex thinking and insight. Therefore, students who scored poorly on standardized tests at an early age would usually be assigned to the remedial or basic skills classes to master those essential basic skills before being exposed to the more challenging and motivating complex thinking skills. In other words, they could not handle the rigor and they went to the "time out" rooms, programs, or grades until they showed they could "merge" back into the regular flow of the general education classes.

Popham (2001) believes that incessant "skill and drill" often turns into "drill and kill." He believes that repetitious instructional activities tend to deaden student's genuine interest in learning. "All the excitement and intellectual vibrancy that students might encounter during a really interesting lesson are driven out by a tedious, test-fostered series of drills" (p. 20). As a result, some students may mentally drop out of schools in the early grades, and physically drop out in high school rather than sit through tedious test-review lessons preparing them to pass high-stakes tests in order to get promoted to the next grade or to graduate from school.

ASSESSMENT AND EVALUATION

Assessment is a global term for gathering information for the purpose of decision making. Chen and McNamee (2007) say, "For classroom teachers, assessment is the process of listening, observing, and gathering evidence to evaluate the learning and development of children in the classroom context" (p. 4). Assessment is the ongoing process of *gathering and analyzing evidence* of what a student can do. *Evaluation* is the process of *interpreting* the evidence and *making judgments* and decisions based on the evidence. If the assessment is not sound, the evaluation will not be sound. Figure 0.1 shows some of the differences between assessment and evaluation, as well as characteristics of performance assessment and portfolios. In most classrooms, teachers assess a student on the basis of observations, oral conversations, and written work. They make instructional decisions based on these assessments. If the assessment is ongoing and

frequent, immediate changes can be made to help the student achieve the desired outcome. If the assessment is flawed, the final evaluation will be based upon invalid and unreliable data. The quality of the final evaluation is only as valid as the quality of the ongoing assessment data upon which it is based.

Four Quadrants of Assessment

Assessment	Evaluation
Assessment "for" Learning • Formative • Ongoing feedback in "real time" • Collection of data • Differentiated to meet student needs • "Do Overs" to allow improvements • Helps teachers improve their teaching • Helps students improve their learning	Assessment "of" Learning • Summative • Final judgment based on evidence • Analysis and evaluation of data • Standardized to test all students • "Last attempt" to meet standards • Used to evaluate teachers' effectiveness • Used to prove the quality of learning to parents, administrators, and policymakers
Performance Assessment	**Student Portfolio**
• Enriched and integrated curriculum • Meaningful and authentic tasks • Real-life applications of knowledge • Student motivation and engagement • Interactive teaching strategies • Differentiated learning options • Collaborative and cooperative learning • Multiple standards addressed • Metacognition and self-assessment	• Collection of evidence • Development and growth • Progress over time • Process used to produce performance • Final product or performance • Reflections on all work • Self-assessments using rubrics • Framework for learning • Examination of student work

Figure 0.1

Diagnostic evaluations are often administered at the beginning of a course, quarter, semester, or year to assess the skills, abilities, interests, levels of achievement, or difficulties of one student or a class. Diagnostic evaluations should be done informally and are not included in the grade. Teachers use the results to modify programs, determine causes of learning difficulties, and ascertain students' learning levels. By having information about the student's entry-level skills, a teacher assesses how far the student has progressed throughout the course or year. Diagnostic assessments are used as baseline data to find out where the students are before a teacher tries a new intervention to produce desired results. Diagnostic tools include items such as pre-tests, writing samples, problem-solving exercises, skill tests, attitude surveys, or questionnaires.

Tomlinson (1999) discusses the strategy of "compacting" in which teachers assess students before beginning a unit of study or development of a skill.

Teachers analyze the results and create a plan to help the students learn the things they are lacking. Students who do well on the pre-assessment should not have to continue to work on what they already know, and teachers can then create enrichment activities to challenge, extend, and motivate these students so they will not get bored. By analyzing the prior knowledge, ability levels, and personal interests of the students at the beginning of a course or school year, teachers are able to differentiate the content, the products, and the process in order to meet their diverse needs. *One size no longer fits all* in the inclusive classroom and it is important to use diagnostic data to determine the academic and social needs of students and then plan subsequent curriculum opportunities and instructional strategies to meet those various needs.

Formative vs. Summative Assessment

According to Popham (2008), the term *formative* was first introduced in an essay about educational evaluation written by Michael Scriven in 1967. Scriven contrasts formative evaluation with summative evaluation as follows:

> If the quality of an early-version educational program is evaluated while the program is still malleable—capable of being improved because of an evaluation's results—this constitutes *formative* evaluation. In contrast, when a mature, final-version educational program is evaluated in order to make a decision about its continuation or termination, this constitutes *summative* evaluation. (as cited in Popham, 2008, p. 3)

The word "malleable" suggests that an educational program is pliable enough to change, adapt, and improve on an ongoing basis. When the term is applied to students, it sounds like teachers are pulling strategies from their bag of ideas to use with each student who is still "a work in progress."

Formative Assessment

The term *formative assessment* has evolved over the past forty years, but the key idea is that teachers should use evidence of learning to *adjust* instruction. Reviews of research in this area were updated by Black and Wiliam (1998) who concluded that "regular use of classroom formative assessment would raise student achievement by 0.4 to 0.7 standard deviations—enough to raise the United States into the top five countries in the international rankings for math achievement, for example" (Wiliam, 2007, p. 189). Formative assessment usually refers to classroom assessments with teachers providing specific feedback to help students. "Measured in terms of impact on student achievement, the single most important thing to change in teachers' practice is the minute-to-minute and day-by-day use of assessment to adjust instruction" (Wiliam, 2007, p. 188).

Formative assessment has been defined as occurring while the learning is taking place and allows for immediate interventions during the process or activity so teachers can modify the teaching and learning activities. Marzano (2006) says:

> By definition, then, formative classroom assessment can and should begin immediately within a learning episode and span its entire duration.

> Additionally, formative classroom assessment can take a wide variety of formats, both formal (e.g., paper-and-pencil quiz) and informal (e.g., discussion with a student). (p. 9)

Formative assessment calls for those "teachable moments" when effective teachers know how to stop in the middle of a lesson and reteach the skill using another strategy if they see their *malleable* students are still confused.

Summative Assessment

Summative assessment is the term used to grade or make final judgments about what the student has learned at the end of instruction; therefore, it usually occurs at the end of a unit, course, or program. It determines the effectiveness or ineffectiveness of already-completed instructional activities. Once teachers collect the assessment information, Airasian (2000) says they use it to make decisions or judgments about pupils, instruction, or classroom climate. He says:

> Evaluation is the process of making judgments about what is good or desirable as in, for example, judging the quality of pupils' essays or the desirability of a particular instructional activity. Evaluation occurs after assessment information has been collected, synthesized, and thought about because that is when the teacher is in a position to make informed judgments. (p. 10)

Summative assessment, then, is the analysis and evaluation of the data that have been collected throughout the formative assessment process. If the data collected are valid and reliable, teachers interpret them and make appropriate judgments about the students and decisions about the programs.

Once the final evaluation is made, it is too late for teachers to adapt instructional activities to meet the needs of their students. It is over! One experienced educator noted that formative assessment helps teachers help *this year's* students succeed whereas summative assessment only shows a teacher *what went wrong.* While it is too late to help this year's students, it might still help the teacher do better with *next year's* students! Unit tests, final research papers, end-of-course tests, final exams, standardized tests, and report cards are generally classified as summative assessments because they represent the end result with no opportunities for students to have another chance to improve and/or meet standards.

DEFINITIONS OF PERFORMANCE ASSESSMENT

Many terms or phrases are used when discussing the alternatives to conventional objective or multiple-choice testing. Alternative assessment, authentic assessment, standards-based assessment, and performance-based assessment are sometimes used synonymously "to mean variants of performance assessments that require students to generate rather than choose a response" (Herman, Aschbacher, & Winters, 1992, p. 2). Performance assessments are

highly engaging for students because they connect their content knowledge with the processes they will use in the real world. Some other characteristics of performance assessments include the following:

- problem scenarios requiring higher-order thinking skills such as analysis, synthesis, and evaluation to solve problems and create original work
- realistic performance tasks correlating to real-life situations faced by students and adults every day
- motivating tasks focusing on producing an authentic product or performance correlated to state standards
- collaboration and group interaction emphasizing both academic and social outcomes
- integration of multiple subject areas highlighting the interdependence of big ideas and essential questions across disciplines

Some other characteristics are included on the performance assessment web in Figure 0.2. Not all of the characteristics need to be present, but they are all important components of effective performance assessment.

Regardless of the different terminologies most of the definitions exhibit two central features: "First, all are viewed as *alternatives* to traditional multiple-choice, standardized achievement tests; second, all refer to *direct* examination of student *performance* on significant tasks that are relevant to life outside of

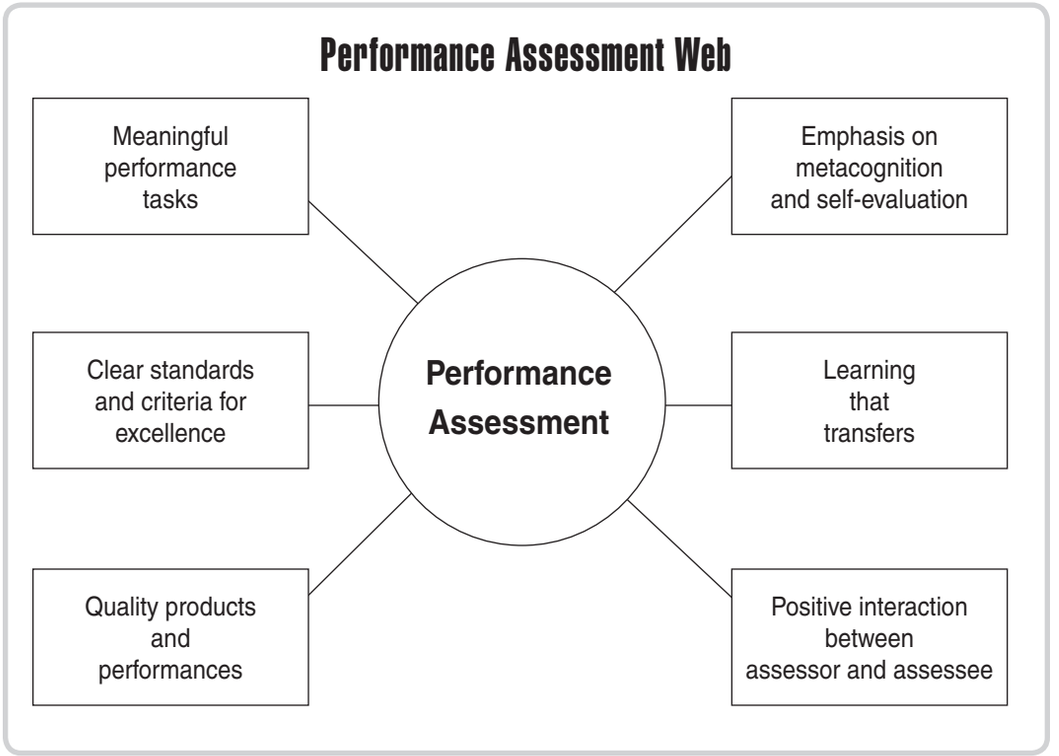

Figure 0.2

school" (Worthen, 1993, p. 445). Almost every state has created "performance standards" for students to demonstrate not only what they know but also how they are able to do things. The final proof is often in the performance.

Archbald and Newmann (1988) describe the term *authentic assessment* as follows: "A valid assessment system [that] provides information about the particular tasks on which students succeed or fail, but more important, it also presents tasks that are worthwhile, significant, and meaningful—in short, *authentic*" (p. 1). Authentic tasks related to real life provide the relevance students need to find meaning in what they are studying. The tasks help students become actively engaged in the learning process, rather than passive receptors of content knowledge.

PORTFOLIOS

A portfolio has been defined as "a *purposeful* integrated collection of student work showing effort, progress, or a degree of proficiency. Portfolios are often defined by the purpose underlying the collection of artifacts and the scope of such purposes is almost unlimited" (Butler & McMunn, 2006, p. 66). Portfolios provide collections of student evidence that show growth and development over time. Portfolios allow students to examine their own work and reflect on their learning. They help students analyze their strengths and weaknesses and set both short- and long-term goals. A portfolio contains both formative and summative evaluations because it is a collection of evidence to show how (or if) students are meeting goals or standards.

Portfolios and e-portfolios include applications of content skills and chronicle students' progress and growth toward meeting curriculum goals and standards. According to Belgrad, Burke, and Fogarty (2008):

> They [portfolios] provide a much richer and more revealing portrait of the student as a learner. Such a picture cannot be captured by a single test score. Within the portfolio process, students become active agents in the acquisition and exposition of their knowledge across the content areas of the grade levels. (p. xv)

One test can be a snapshot of a student on one day, but it does not show the "whole child" or how the student progresses and grows over time. Portfolios also show the process students used to achieve their final products. Many portfolios contain first, second, and third drafts along with the final research paper in order to show the pathway toward improvement.

BALANCED ASSESSMENT

Assessment should not have to generate an "either/or" or a "throw out the baby with the bath water" approach. Most educators agree with Stiggins that we need all the tools at our disposal. Shulman (1988) talks about teacher assessment suggesting that no one piece of evidence is sufficient to provide an

evaluation. Teachers need to combine various methods of assessment so that the strengths of one offset the limitations of the other.

Student assessment should follow the same guidelines. No one assessment tool by itself is capable of producing the quality information needed to make an accurate judgment of a student's knowledge, skills, understanding of curriculum, motivation, social skills, processing skills, and lifelong-learning skills. Each single measurement by itself is insufficient to provide a true portrait of the student or learner. If educators combine standardized and teacher-made tests measuring knowledge and content with portfolios measuring process and growth, and performances measuring application, they will provide a more accurate portrait of the individual learner. Fogarty and Stoehr (2007) discuss the balanced assessment model (see Figure 0.3) used to address traditional assessments, portfolio assessments, and performance assessments and subsequently meet the needs of all students.

AND NOW . . . THE TOOLS!

Performance assessments provide teachers with a repertoire—a vast array of tools to measure student growth. The following chapters focus on specific tools teachers need to create a vivid, colorful, and true portrait of students as they develop and grow over the course of a year. In the past, progress was chronicled by a superficial "snapshot" of the student. The snapshot usually consisted of a few pictures of standardized test scores, midterms, final grades, and other one-dimensional scores that lay lifeless in the permanent record file. The grades on the report cards do not adequately describe the skills the students had when they entered a class, as compared to the skills they had when they left the class. Nothing more than a static glimpse of a student can be gleaned from the traditional, cumulative record system that has dominated our school systems for the past two centuries. Teachers who create a repertoire of assessment tools allow students to show what they know and what they can do in multiple formats. Each tool represents a different way to challenge students to succeed and to measure their progress along the way.

Each chapter in this book will introduce an assessment tool to record a student's growth and achievement. The chapters include a description of *what* is the tool, *why* we should use the tool, and *how* we could use the tool. Examples of many of the assessments are provided, and teachers will have a chance to create original tools at the end of each chapter in the "On Your Own" section, as well as self-evaluate their work on the "Reflection Page." The fifth edition of *How to Assess Authentic Learning* presents only a few of the many options available for teachers to add to their repertoire of assessment strategies. As educators review the strategies in this book, they should take some time to reflect on how the tools provide a variety of strategies to help all students meet and exceed the standards.