Considering Evidence of Learning in Diverse Classrooms

What should count as evidence of learning? Of understanding?
How might we differentiate our assessments without sacrificing validity and reliability?
How can we maintain standards without standardization?
How can assessment promote learning, not simply measure it?

Anyone concerned about teaching and learning is automatically interested in assessment. Assessment provides us with evidence to help answer important questions: "Did the student learn it?" "To what extent does the student understand?" "How might I adjust my teaching to be more effective for learners with varying needs?" The logic of backward design signals the importance of "thinking like an assessor" by placing Stage 2 (determining acceptable evidence) *before* Stage 3 (planning teaching and learning activities). By considering in advance the assessment evidence needed to validate that the desired results have been achieved, teaching becomes more purposeful and focused. Also, with clarity about what constitutes evidence that students have achieved desired results, teachers have a consistent framework within which they can make modifications for their students' readiness levels, interests, and learning preferences.

Principles of Effective Assessment

Three key principles should inform and guide classroom assessment. We'll now explore their conceptual foundation and consider the practical applications of each within academically diverse classrooms. Each of the

principles provides a rationale for attending to student variance within the parameters of best practice.

Assessment Principle 1: Consider Photo Albums Versus Snapshots

Assessment is a process by which we make inferences about what students know, understand, and can do based on information obtained through assessment. Although educators sometimes loosely refer to an assessment as being valid and reliable, in fact a more precise conception has to do with the extent to which the results of an assessment permit valid and reliable inferences. Because all forms of assessment have inherent measurement error, our inferences are more dependable when we consider more than one measure. In other words, reliable assessment demands multiple sources of evidence.

Consider this principle in terms of a photographic analogy. A photo album typically contains a number of pictures taken over time in different contexts. When viewed as a whole, the album presents a more accurate and revealing "portrait" of an individual than does any single snapshot within. It is the same with classroom assessment—a single test at the end of instruction is less likely to provide a complete picture of a student's learning than a collection of diverse sources of evidence is.

Professional measurement specialists (psychometricians) understand this basic assessment principle. For example, Dr. Michael Kean (1994), vice president for CTB/McGraw-Hill, a major publisher of standardized tests, states: "Multiple measures are essential because no one test can do it all. Therefore, no test, no matter how good it is, should be the sole criterion for any decision."

Unfortunately, most politically driven accountability systems in North America rely on "quick and dirty" standardized tests (which provide a snapshot rather than a photo album) as a basis for judging students, schools, and districts. There is nothing inherently wrong with standardized tests. They provide useful and comparable data about student achievement levels on easily tested content goals. However, the problem occurs when the results of a *single* test are used to make high-stakes decisions. The widespread use of one-shot accountability testing has consequences that are well documented and include the following:

- The pressures to improve test scores can lead to a narrowing of the curriculum toward the tested topics and an overemphasis on "test prep" at the expense of meaningful learning.
- Important educational goals that are not easily and cheaply tested in a large-scale context (e.g., oral communication, decision making, research, expression in the arts) can fall through the cracks if they are not measured.
- The standardized nature of most large-scale, "one-size-fits-all" testing flies in the face of what we know (i.e., not every child learns in the same way at the same time).
- The predominant assessment format (selected-response) favors students with facility for recall and recognition. The results of high-pressure exams in which reading ability is paramount may present a distorted picture of the achievement of learners whose parents do not speak standard English, as well as of students with disabilities.

The overreliance on a single measure as a basis for inferences and high-stakes decisions is psychometrically unsound and politically risky, as recent testing scandals attest (Hendrie, 2002; White, 1999). However, our focus is less about the ills of accountability testing than it is about those aspects of assessment that we can influence—the assessments that we use in our classrooms, schools, and districts.

It is in the context of the classroom that the application of "assessment as photo album" is most feasible and natural. Indeed, a variety of classroom assessments may be used to gather evidence of learning (McTighe & Wiggins, 2004):

- Selected-response format (e.g., multiple-choice, true-false) quizzes and tests
- Written or oral responses to academic prompts (short-answer format)
- Performance assessment tasks, yielding
  - Extended written products (e.g., essays, lab reports)
  - Visual products (e.g., PowerPoint shows, murals)
  - Oral performances (e.g., oral reports, foreign-language dialogues)
  - Demonstrations (e.g., skill performances in physical education)

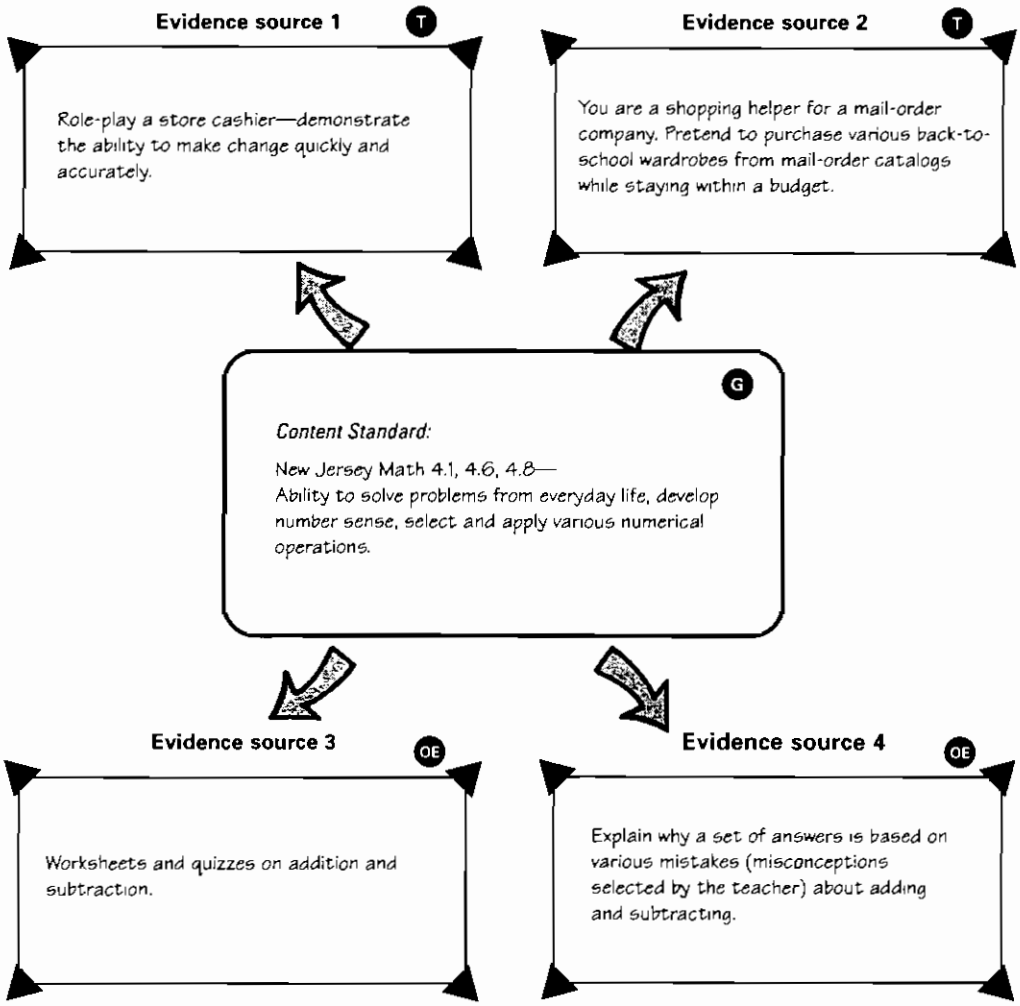
- Long-term, "authentic" projects (e.g., senior exhibitions)
- Portfolios (systematic collections of student work over time)
- Reflective journals or learning logs
- Informal, ongoing observations of students (e.g., teacher note taking, probing questions, exit cards, Quick-Writes)
- Formal observations of students using observable indicators or criterion list
- Student self-assessments
- Peer reviews and peer response groups

In planning for classroom assessments, consider the "photo album" graphic organizer in Figure 5.1. This figure illustrates the use of multiple sources of evidence for assessing attainment of an important content standard, in this case arithmetic problem solving. Although we are not suggesting that *everything* we teach requires multiple assessments, we do believe that more than a single source of evidence is needed for our significant, "essential and enduring" goals. That recommendation may mean providing more than one format option for key assessments. It will certainly mean varying format options over the course of a unit of study. Both are clearly important in academically diverse classrooms where different students will most fully be able to demonstrate their knowledge, understanding, and skill in different formats.

Although useful for individual planning, this graphic organizer has proven especially valuable for team planning. As suggested in Chapter 3, the logic of backward design dictates that evidence of learning (Stage 2) must be derived from the desired results (Stage 1), and this logic applies to teachers working in grade-level and department teams as well. In standards-based education, the rubber meets the road with assessment. Unless we agree not only on the goals but also on the needed assessment evidence of meeting them, we cannot claim that our teaching is standards based. By working with colleagues to forge consensus about what it looks like when students achieve desired results, educators realize more coherent curricula, more reliable assessments, and greater consistency in grading and reporting *across* classrooms and schools.

Including a variety of assessments is important not only from a measurement perspective but as a matter of sensitivity to varied learners. Because students differ in their preferred way of showing what they have learned, providing multiple and various assessment types increases the opportunity for students to work to their strengths and, ultimately, the likelihood of their success. Like the judicial system, we need a "preponderance of evidence" to convict students of learning! Ultimately, the validity and reliability of our

FIGURE 5.1
An Assessment Photo Album



Source: From *Understanding by Design Professional Development Workbook* (p. 146), by J. McTighe and G. Wiggins, 2004, Alexandria, VA: Association for Supervision and Curriculum Development. Copyright 2004 by the Association for Supervision and Curriculum Development. Adopted with permission.

judgments about student achievement are enhanced when we ensure that the types of assessment we use are effective for particular learners in providing evidence of their achievement.

Assessment Principle 2: Match the Measures with the Goals

To allow valid inferences to be drawn from the results, an assessment must provide an appropriate measure of a given goal. Thus, thinking about assessment evidence in Stage 2 cannot be done without a careful consideration of the desired results (Stage 1). We have found it useful to distinguish among three types of educational goals: (1) *declarative knowledge*—what students should know and understand, (2) *procedural knowledge*—what students should be able to do, and (3) *dispositions*—what attitudes or habits of mind students should display (Marzano, 1992). These categories have direct implications for how we teach and assess. For example, if we want to see whether students know multiplication tables or chemical symbols *(declarative knowledge)*, then objective test items, such as multiple-choice, matching, true-false, or fill-in-the-blank, will provide the appropriate evidence in an efficient manner. When we wish to check for proficiency in skill/process areas such as drawing, writing, or driving *(procedural knowledge)*, some type of performance assessment is needed. For *dispositions*, such as "appreciation of the arts" or "persistence," evidence will have to be collected over time through observations, examples, portfolios, and self-assessments. After all, a quiz on "persistence" would be an inappropriate measure of such a goal.

In a differentiated classroom, there is particular meaning in attending to student proficiency with all three kinds of knowledge. Some students will need additional support, for example, with the procedural knowledge (skills) in a unit but be progressing well with the declarative (knowledge and understanding), whereas others will exhibit the reverse profile at a given time. If a teacher is to use assessment data to map instructional plans, it matters that the data provide information on student strengths and needs with essential knowledge, understanding, and skill. Without using such individual-specific data, we give ourselves permission to teach in a one-size-fits-all fashion—asking the impossible of some students while teaching others what they already know (Taba & Elkins, 1966). In addition, data on student dispositions or habits of mind and work can yield important insights about why a particular student is (or is not) progressing at a given time. Furthermore, data

on student dispositions becomes important in reporting student progress in a differentiated classroom. This topic will be discussed further in Chapter 8.

Diversity of goals implies that we should include a variety of assessment pictures in our assessment "photo album." We accomplish this by selecting various assessment formats to give us appropriate measures for our goals. Yet, despite the importance of collecting multiple pieces of evidence and matching the measures with goals, we often observe teachers making assessment decisions based on what assessment is easiest to give and grade. This is understandable given the time- and labor-intensiveness of some types of assessment and the pressures to "defend" grades to students, parents, and administrators. Nevertheless, we strongly recommend that our goals should dictate the nature of our assessments, not external factors. It is incumbent upon school and district leaders to establish structures (e.g., time for group scoring of student work and realistic report card completion timelines) so that responsible assessment practices can be enacted feasibly.

Assessing Understanding

In Chapter 3, we discussed the value of identifying the "big ideas" that we want students to come to understand. Now, we'll take a finer-grained look at this particular goal of "understanding" by examining three questions: What is the difference between knowing and understanding? How will we know that students *truly* understand the big ideas that we have identified? How might we allow students to demonstrate their understanding in diverse ways without compromising standards?

Knowing is binary—you either know something or you don't. Declarative knowledge of facts and basic concepts falls into this category, and assessing such factual knowledge can be readily accomplished through objective tests and quizzes featuring "correct" answers. Understanding is more a matter of degree, as our language suggests. For example, we speak of someone having a sophisticated insight, a solid grasp, an incomplete or naive conception, or a misunderstanding. Thus, when we ask, "To what extent does she understand?" the answer is revealed along a continuum as shades of gray, rather than black and white. This point has implications for how we assess and how we describe the results.

A challenge for assessing understanding is found in the word itself—understand has different connotations. For instance, consider these four uses of the term:

- They really *understand* Spanish.
- She *understands* what I am going through.
- He knows the historical facts but does not *understand* their significance for today.
- I now *understand* that I never saw the big picture.

The first example suggests that understanding a language enables someone to *use* it—that is, to communicate effectively via listening, speaking, reading, and writing. In the second example, the emphasis is on *empathy*—the capacity to feel as someone else. The third case implies *transfer*—the ability to apply what one has learned in a new situation. The fourth example is *metacognitive;* that is, the individual is capable of reflecting on his or her thinking and learning processes.

The fact that the term *understand* can be used in such diverse ways has led some researchers and educators to decry its use in framing goals. They argue that it is too ambiguous to provide goal clarity and measurement specificity.

Taking a different tack, Wiggins and McTighe (1998, 2005) propose that these various connotations can be used to formulate a conception of understanding for assessment purposes. They propose that understanding is revealed through six facets, summarized in Figure 5.2.

These six facets do not present a theory of how people come to understand something. We'll leave that to the cognitive psychologists to explain. Instead, the facets are intended to serve as *indicators* of how understanding is revealed. Thus, they provide guidance as to the kinds of assessments we need to determine the extent of student understanding.

Although the six facets offer a full array of possible indicators of understanding, a basic approach for determining whether learners really understand involves two: explain and apply. When we speak of explanation, we seek more than a memorized recitation. Doctoral students are required to *defend* their dissertation in order to demonstrate to their committee that they understand their research and its meaning. The same idea applies, albeit in a less formalized manner, when we ask learners to "put it in their own words,"

FIGURE 5.2
The Six Facets of Understanding

When we truly understand, we

- Can **explain** via generalizations or principles: provide justified ond systematic accounts of phenomena, facts, and data; make insightful connections and provide illuminating examples or illustrations.
- Can **interpret:** tell meaningful stories; offer apt translations; provide a revealing historical or personal dimension to ideas and events; make it personal or accessible through images, anecdotes, onalogies, and models.
- Can **apply:** effectively use and adapt what we know in diverse and real contexts—we can "do" the subject.
- Have **perspective:** see and hear points of view through critical eyes and ears; see the big picture.
- Display **empathy:** find value in what others might find odd, alien, or implausible; perceive sensitively on the basis of prior direct experience.
- Have **self-knowledge:** show metacognitive awareness; perceive the personal style, prejudices, projections, and habits of mind that both shape and impede our own understanding; be aware of what we do not understand; reflect on the meaning of learning and experience.

give reasons for their answers, support their position, justify their solution, and show their work.

It is important to note that explanations need not be exclusively verbal (written or oral). Visual explanations in the form of concept maps, sequence chains, flowcharts, visual analogies, and so on, can be quite revealing and may be particularly beneficial in ensuring that students who have strong visual preferences or who struggle with verbal expression have an opportunity to express what they are learning.

When we call for application, we do not mean a mechanical response or mindless "plug-in" of a memorized formula. Rather, we ask students to transfer—to use what they know in a new situation. We recommend that teachers set up realistic, authentic contexts for assessment; when students are able to apply their learning thoughtfully and flexibly, true understanding is demonstrated. Consider an analogy here. In team sports, coaches routinely conduct drills to develop and refine basic skills. However, these practice

drills are always purposefully pointed toward performance in the game. Too often, we find that classrooms overemphasize decontextualized drills and provide too few opportunities for students to actually "play the game." Figure 5.3 differentiates between inauthentic drills and authentic application.

FIGURE 5.3
Inauthentic Versus Authentic Work

Inauthentic Work	Authentic Work
Fill in the blank	Conduct research using primary sources
Select an answer from given choices	Debate a controversial issue
Answer recall questions at end of chapter	Conduct a scientific investigation
Solve contrived problems	Solve "real-world" problems
Practice decontextualized skills	Interpret literature
Diagram sentences	Do purposeful writing for an audience

Both drills and authentic application are necessary in the field and the classroom. Students need to master the basics, and skill drills support that need. But learners also need a chance to use their knowledge and skills—in other words, to "do" the subject.

When students can apply knowledge and skill appropriately to a new situation and can effectively explain *how* and *why*, we have the evidence to "convict" them of understanding.

Let's consider two examples of assessment tasks that require application and explanation for a middle grades unit on nutrition.

- Because our class has been learning about nutrition, 2nd grade teachers in our school have asked for our help in teaching their students about good eating. Create an illustrated brochure to teach the 2nd graders about the importance of good nutrition for healthful living. Use cut-out pictures of food and original drawings to show the difference between a balanced

diet and an unhealthy diet. Show at least two health problems that can occur as a result of poor eating. Your brochure should also contain accurate information and should be easy for 2nd graders to read and understand.

- Because we have been learning about nutrition, the camp director at the Outdoor Educational Center has asked us to propose a nutritionally balanced menu for our three-day trip to the center later this year. Using the USDA Food Pyramid guidelines and the nutrition facts on food labels, design a plan for three days, including the three main meals and three snacks (morning, afternoon, and campfire). Your goal: a healthy and tasty menu. In addition to your menu, prepare a letter to the director explaining how your menu meets the USDA nutritional guidelines. Include a chart showing a breakdown of the fat, protein, and carbohydrate content and vitamins, minerals, and calories. Finally, explain how you have tried to make your menu tasty enough for your fellow students to want to eat.

Notice that in both examples, students are asked to apply their knowledge of nutrition to a real-world situation and include an explanation. They are required to use what they know in flexible ways to meet a goal for an identified audience. Both tasks are open-ended in that they allow students to personalize their response while still meeting established criteria—an example of standards without standardization. Such assessments provide evidence of meaningful learning in a qualitatively different way than would an objective test of nutrition facts (although we might well include such a test as part of our photo album). Certainly, in a differentiated classroom a teacher acknowledges that although it is not negotiable that a student demonstrate understanding, how that student might best do so *is* highly flexible. Furthermore, it is quite possible that some students will be appropriately challenged by an assessment task that is more complex and requires more advanced manipulation of skills, whereas other students need a task that is more concrete and requires a more fundamental, foundational, or familiar application of skills. That the students must show understanding of essential big ideas does not vary, but the "degree of difficulty" of the assessment task can vary to appropriately address variety in learner readiness.

The GRASPS Frame

As a means of creating more authentic "performances of understanding," we recommend that teachers frame assessment tasks with the features suggested by the acronym GRASPS. In other words, include (1) a real-world **goal,** (2) a meaningful **role** for the student, (3) authentic (or simulated) real-world **audience(s),** (4) a contextualized **situation** that involve real-world application, (5) student-generated culminating **products** and performances, and (6) consensus-driven performance **standards** (criteria) for judging success. Notice these elements in the two previously presented examples.

We do not mean to imply that *everything* we teach or assess needs to be framed using GRASPS. However, for those important ideas and processes that you really want students to understand, we believe that more authentic tasks have merit. Performance tasks having these features provide meaningful learning targets for learners, worthy performance goals for teaching, and the kind of evidence needed to assess true understanding.

Moreover, it is important to stress that virtually all students in our schools should have regular opportunities to demonstrate their proficiency with important content goals through assessments that embody the GRASPS characteristics. Some parameters for student work and teacher scaffolding of student success may well need to vary among students, but not the opportunity to express learning through meaningful assessments that include student choice, that are focused on essential content goals, and that are judged according to substantive criteria. A highly advanced learner, for example, may apply understandings in a less familiar or less well-defined context or for an audience with sophisticated knowledge of the domain in question. A student who struggles to learn may apply understandings in a more familiar or more structured context or for an audience of peers or younger students. Both students should be expected to demonstrate genuine understanding of essential principles in real-world situations.

Assessment Principle 3: Form Follows Function

The way in which we design and use classroom assessments should be directly influenced by the answers to four questions: What are we assessing? Why are we assessing? For whom are the results intended? How will the results be used? We have discussed the relationship between what and how we assess in the previous section. Now we turn our attention to purpose.

Classroom assessments serve different purposes, one of which is summative. *Summative* assessments are generally used to summarize what has been learned. These assessments tend to be evaluative in nature, and their results are often encapsulated and reported as a score or a grade. Familiar examples of summative assessments include tests, performance tasks, final exams, culminating projects, and work portfolios. These evaluative assessments command the attention of students and parents, because their results typically "count" and become recorded on report cards and transcripts.

In addition to evaluation, two other assessment purposes—diagnostic and formative—are critical to teaching and learning. *Diagnostic* assessments (or pre-assessments) typically precede instruction and are used to check students' prior knowledge and skill levels and identify misconceptions, interests, or learning style preferences. They provide information to assist teacher planning and guide differentiated instruction. Examples of diagnostic assessments include skill checks, knowledge surveys, nongraded pre-tests, interest or learning preference checks, and checks for misconceptions.

Formative assessments occur concurrently with instruction. These ongoing assessments provide information to guide teaching and learning for improving achievement. Formative assessments include both formal and informal methods, such as ungraded quizzes, oral questioning, observations, draft work, think-alouds, student-constructed concept maps, dress rehearsals, peer response groups, and portfolio reviews.

Although summative/evaluative assessments often receive the most attention, diagnostic and formative assessments provide critical "along the way" information to guide instruction in response to the nature and needs of the diverse learners. Waiting until the end of teaching to find out how well students have learned is simply too late. Just as the most successful coaches and sponsors of extracurricular activities such as yearbook, orchestra, theater, and athletics recognize the importance of ongoing assessments and continuous adjustments as the means to achieve maximum performance, so do the best teachers. As a validation of good instincts, recent research has confirmed the benefits of regular use of diagnostic and formative assessments as feedback for learning (Black & William, 1998). In a differentiated classroom, a teacher continuously examines ongoing assessment data for individuals as a means of adapting "up-front" teaching plans so that they address particular learner needs. As noted educator Hilda Taba pointed out,

"Diagnosis, of course, is never completed. Every contact with students reveals something that the teacher did not know before, something important for intelligent planning of instruction" (Taba & Elkins, 1966, p. 24).

Responsive Assessment to Promote Learning in Diverse Classrooms

We conclude this chapter by describing four classroom assessment practices that honor student differences and promote learning.

Assess *Before* Teaching

Diagnostic assessment (pre-assessment) is as important to teaching as a physical exam is to prescribing appropriate medical regimens. At the outset of any unit of study, some students are likely already to have mastered many of the skills that the teacher is about to "introduce," and they may already have a relatively sophisticated understanding of some or all of the unit's enduring understandings. Simultaneously, some students are likely to be deficient in precursor skills necessary to become proficient with the unit's essential skills and to lack a context or experience base for beginning a study of the unit's enduring understandings. A teacher who intends to support success for each learner needs a sense of the learners' starting points as a unit begins. "Teaching in the dark is questionable practice" (Taba & Elkins, 1966).

Pre-assessments should focus on the unit's essential knowledge, understanding, and skill. They should provide a window into important strengths and weaknesses that students may bring to the study. Furthermore, they should not be graded. Rather, pre-assessments contribute to a teacher's general sense of each student's readiness status relative to essential content goals for the unit. At key points in the year, pre-assessments may also be useful in gaining insights about a student's interests or preferred routes to learning. Many formats are useful for pre-assessment, including 3-2-1 cards, Frayer diagrams, quizzes, journal entries, checklists, and concept maps.

Informed with a sense of students' varying learning needs, a teacher can begin to form instructional groups, assign appropriate student tasks, locate appropriate learning materials, and so on. Then, throughout the unit,

formative assessments continue to assist the teacher in refining his or her understanding of a learner's needs and in responding to learners in ways likely to maximize their growth.

Offer Appropriate Choices

Responsiveness in assessment is as important as it is in teaching. Just as students differ in their preferred ways of taking in and processing information, so do they vary in the manner by which they best show what they have learned. Some students need to "do," whereas others thrive on oral explanations. Some excel at visual representations; others are adept at writing. To make valid inferences about learning, teachers need to allow students to work to their strengths. A totally standardized, one-size-fits-all approach to classroom assessment may be efficient, but it is not "fair," because any chosen format will favor some students and penalize others.

Assessment becomes responsive when students are given appropriate options for demonstrating knowledge, skill, and understanding. In other words, allow some choices—but always with the intent of collecting *needed evidence based on goals*. Without a clear connection between the desired results and the required evidence, teachers will be stuck assessing apples, oranges, and grapes.

An adaptation of tic-tac-toe provides a structure for giving students choices of products and performances while keeping the end in mind. Figure 5.4 illustrates one example in which the teacher structures product and performance options of various genres by which students could display their content understanding.

The tic-tac-toe format enables teachers to structure the options while giving the students choices. The choice options are flexible. For example, if we want students to write, then we would ask all learners to choose one option from the first column and then one other product/performance from the second or third columns. If we seek an accurate and complete explanation, we might give students greater freedom to choose options from the other columns. Figure 5.5 shows a tic-tac-toe chart with greater openness. The "FREE" blocks allow students to propose an alternative source of evidence that suits their strength. For a major project, we might allow students to produce three products, picking one from each column.

FIGURE 5.4
Product and Performance Tic-Tac-Toe (Version 1)

Written	Visual	Oral
Research report	Poster	Lesson presentation
News article	Graphic organizer	Oral presentation
Information brochure	PowerPoint	Radio interview

FIGURE 5.5
Product and Performance Tic-Tac-Toe (Version 2)

Written	Visual	Oral
FREE	Poster	Speech
Persuasive essay	FREE	Debate
Editorial	Campaign poster	FREE

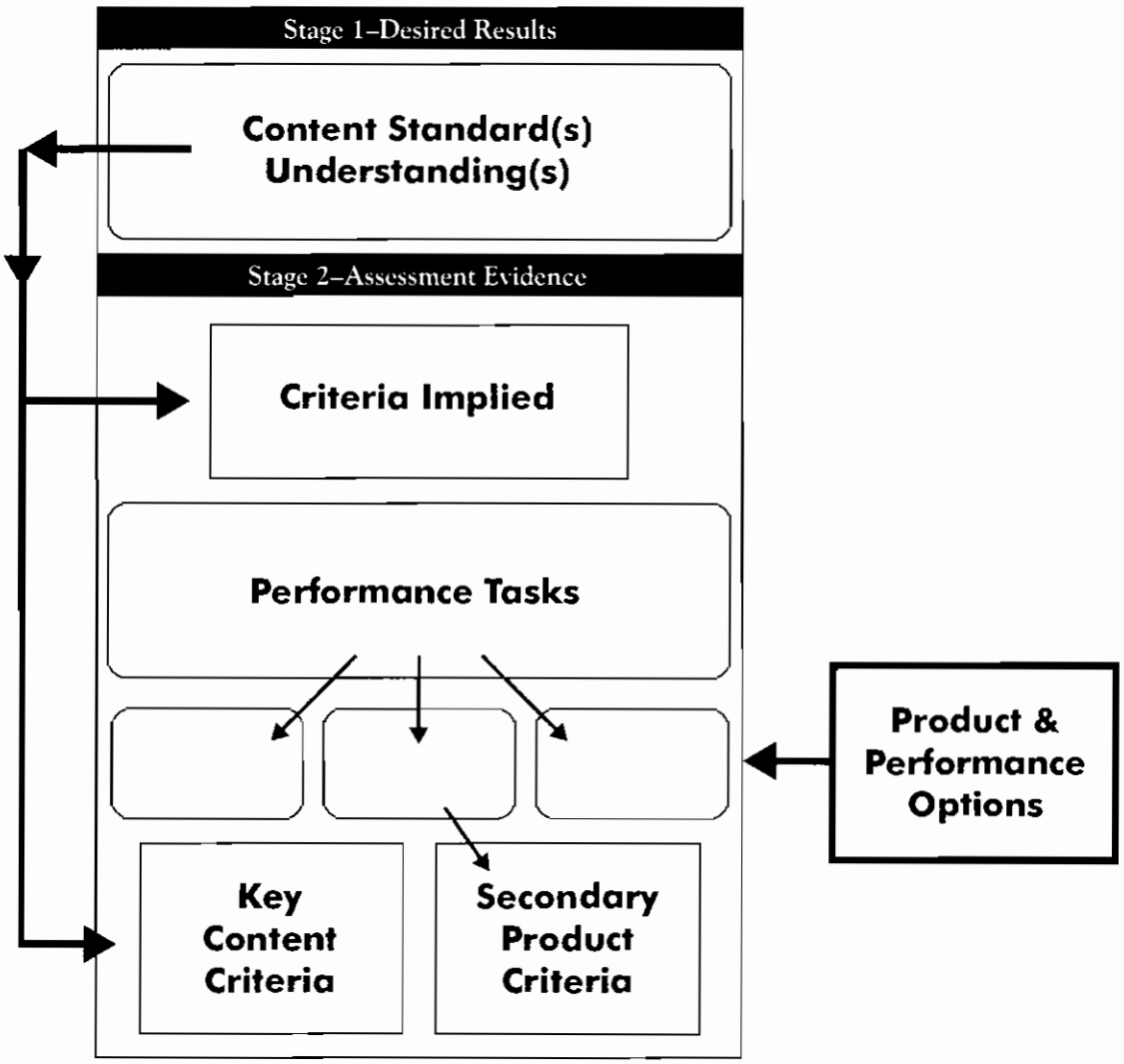
Regardless of how open-ended the task and how many product/performance options are provided, it is imperative that we identify a *common* set of evaluative criteria. This advice might seem counterintuitive; that is, how can we have the *same* criteria if we give students *different* product options? The answer goes back to the logic of UbD's backward design. The general assessment evidence we need to collect in Stage 2 is determined by the desired results identified in Stage 1. However the particulars of an assessment task may be structured so as to allow student choice as discussed earlier. For example, within a unit on nutrition we want students to show their understanding of a balanced diet. This understanding could be assessed by a task that asks students to explain the concept and offer an illustrative example, and the needed evidence could be obtained in writing, orally, or visually. However, regardless of the response mode, all students would be judged by a rubric containing the following key criteria connected to the content: *clear, accurate,* and *complete explanation* of "balanced diet," with an *appropriate example* that *illustrates* the concept. In other words, the criteria are derived primarily from the content goal, not the response mode.

Now, we may wish to add student-specific criteria for the needs of particular learners. For instance, a teacher may stress the use of primary resources in research work undertaken by a highly able 4th grader, whereas secondary sources are appropriate for other learners in the class. (This illustration assumes that use of primary sources is *not* a content goal for the unit.) Likewise, a teacher may add product-specific criteria for the different product genres. For example, if a student prepares a poster to illustrate a balanced diet, we could look for *neatness, composition,* and effective use of *color.* Likewise, if a student made an oral presentation, we could judge her *pronunciation, delivery rate,* and *eye contact* with the audience. However, in this example we consider these to be secondary criteria linked to specific products/performances, rather than the key criteria determined by the content goal. (Note that a speech teacher would use the last set of criteria as key because of their importance to the content standard of effective speaking.)

Of course, we want students to do quality work, regardless of what options they select. But more important, we need to employ the criteria called for by the content goals. If we vary *these* key criteria for different students based on the products they select, then we no longer have a valid and reliable assessment measure. See Figure 5.6 for a visual summary of these points.

We conclude this section with three cautions. First, we must always keep in mind that our aim is to collect appropriate evidence of learning based on the goals, not to simply offer a "cool" menu of product possibilities. If a content standard calls for proficiency in writing or oral presentation, it would be inappropriate to provide alternative performance options other than writing or speaking—except in cases of students for whom writing or speaking is inordinately difficult because of disabilities.

FIGURE 5.6
Criteria and Differentiated Assessments



Second, the options we provide must be worth the time and energy required. It would be inefficient to have students develop an elaborate three-dimensional display or an animated PowerPoint show for content that could be efficiently and appropriately assessed with a multiple-choice quiz. In the folksy words of a teacher friend, "The juice must be worth the squeeze."

Third, feasibility must be considered. Ideally, we might wish to individualize all major assignments and performance assessments, but realistically we have only so much time and energy. Therefore, educators must be judicious in determining when it is important to offer product and performance options (and how many should be offered), striking a balance between a single path and a maze of options.

Despite the challenges, we believe that efforts to provide options for assessment are well worth the trouble. Students given appropriate choices on respectful tasks are more likely to put forth effort and feel a genuine sense of accomplishment for a job well done.

Provide Feedback Early and Often

Legendary football coach Vince Lombardi summed it up: "Feedback is the breakfast of champions." All types of learning, whether on the practice field or in the classroom, require feedback. Ironically, the high-quality feedback systems necessary to enhance learning are limited in our schools, at least in academic classrooms. Consider the observations of assessment expert Grant Wiggins (1998):

> If I had to summarize what I have seen over the past decade in all kinds of schools (public and private; elementary, secondary and collegiate; with and without state testing programs), I would have to report that many educators seem to believe that feedback means giving lots of approval, and some disapproval and advice. In classrooms, the most common piece of so-called feedback is "Good job!" or some equivalent phrase.
>
> It is, of course, important to praise students because it often satisfies and encourages them, but it cannot help them to improve their performance. Praise keeps you in the game; real feedback helps you get better. Feedback tells you what you did or did not do and enables you to self adjust. Indeed, the more self-evident feedback, the more autonomy the performer develops, and vice-versa. (p. 46)

Four qualities characterize an effective feedback system. The feedback must (1) be timely, (2) be specific, (3) be understandable to the receiver, and (4) allow for adjustment. Waiting three weeks (or three months) to find out

how you did on a standardized test will not help your learning. Learners need to find out promptly their strengths and weaknesses in order to improve. The greater the delay, the less likely it is that the feedback will be helpful or used.

Not surprisingly, the best feedback is often observed in the "performance-based" subjects such as art, music, drama, speech, vocational and technical education, family and consumer sciences, and physical education. We also see feedback effectively employed in athletics and extracurriculars, such as band, newspaper, and debate. Indeed, the essence of "coaching" involves ongoing assessment and feedback, as Lombardi's comment suggests.

Specificity is key to focused adjustment. Too many educators consider grades and scores as "feedback," when, in fact, they fail the specificity test. Pinning a letter (*B*-) or a number (82%) on a student's work is no more helpful than comments such as "Way to go" or "Try harder." Although good grades and positive remarks may feel good, they do not advance learning. Specific feedback sounds different—for example, "Your research paper is well organized and contains lots of specific information. You used multiple sources and documented them appropriately. However, your paper lacks a clear conclusion, and you never answered your basic research question." From this feedback, the report writer knows specifically where the paper is strong and what revisions are needed.

Because feedback is directed to the learner, it must be understood. Rubrics are often viewed as feedback tools and can indeed serve in this capacity. However, sometimes the language in a rubric can be lost on a student. Exactly what does the teacher mean by "elegant reasoning" or "sophisticated analysis"? If we want feedback to inform learners and guide their improvement, our feedback must be clear and comprehensible. One approach is to develop "kid language" rubrics. For instance, instead of "documents the reasoning process," we might say, "Show your work in a step-by-step manner so that others can see how you were thinking."

A second approach for making feedback understandable involves the use of models and exemplars. Experienced teachers have a clear conception of what we mean by "well organized," but there is no guarantee that the phrase will convey that same idea to students. They are more likely to understand our feedback when we show several examples that are well organized and easy to grasp compared with several that lack organization and are difficult to follow. If we expect students to act on our feedback, they have to under-

stand it. The use of models helps to make the "invisible visible" through tangible examples. It is also possible to share with individuals or small groups of students exemplars of work completed by students who were at their approximate level of proficiency and who did—or did not—demonstrate proficiency in their work. In that way, students can see work that "looks like they might have done it" and simultaneously see examples of next steps in quality that they believe they could achieve with effort and support.

Here's a simple, straightforward test for a feedback system: Can the learners tell *specifically* from the given feedback what they have done well and what they could do next time to improve? If not, the feedback is not yet specific or understandable enough for the learner.

Finally, the learner needs opportunities to act on the feedback—to refine, revise, practice, and retry. Writers rarely compose a perfect manuscript on the first try, which is why the writing process stresses cycles of drafting, feedback (from self-assessment, peer review, and teacher comments), and revision as the route to excellence. The same process applies in any subject where the goal is deep understanding and fluent performance. Therefore, teachers should build into their instructional plans regular opportunities for feedback and refinement. Learning demands it.

Encourage Self-Assessment and Reflection

The most effective learners are metacognitive; that is, they are mindful of how they learn, set personal learning goals, regularly self-assess and adjust their performance, and use productive strategies to assist their learning. Less effective learners seem to go through school as if in a cloud. They seem clueless about their preferred learning style and about strategies that can enhance their achievement.

Research and experience have shown that metacognitive strategies can be taught, and the benefits to learners can be noteworthy (Bransford, Brown, & Cocking, 2000; Costa & Kallick, 2000; Flavell, 1985). One straightforward approach to cultivating metacognition involves having learners regularly respond to reflective questions such as those listed here (McTighe & Wiggins, 2004). Such questions encourage students to reflect on their learning, consider transfer possibilities, self-assess their performance, and set goals:

What do you really understand about _____?

What questions/uncertainties do you still have about _____?

What was most effective in _____?

What was least effective in _____?

How could you improve_____?

What would you do differently next time?

What are you most proud of?

What are you most disappointed in?

How difficult was _____ for you?

What are your strengths in _____ ?

What are your deficiencies in _____ ?

To what extent has your performance improved over time?

How does your preferred learning style influence _____ ?

What grade/score do you deserve? Why?

How does what you've learned connect to other learning?

How has what you've learned changed your thinking?

How does what you've learned relate to the present and future?

What follow-up work is needed?

Such self-assessment in a differentiated classroom also enables student and teacher to focus both on nonnegotiable goals for the class and personal or individual goals that are important for the development of each learner. For instance, if students had opportunities to reflect on the appropriateness of the degree of task difficulty for them, to name their particular strengths and weaknesses, to think about how their learning preferences work for and against them, and to set personal improvement goals, it is likely that they would have more ownership in both their learning and their classroom where the teacher works to understand and respond to their needs.

Another simple yet effective strategy for providing feedback while encouraging self-assessment and goal setting is to adjust the format of a rubric. Notice in Figure 5.7 that two small squares have been inserted in the bottom left and right corners of each box in an analytic rubric. The squares on the left side enable students to self-assess their performance according to the established criteria and performance levels *before* they turn in their work. The teacher then uses the right-side squares to evaluate. Ideally, the

two judgments would be close. If not, the discrepancy raises an opportunity to discuss the criteria, expectations, and performance standards. Over time, teacher and student judgments tend to converge; in fact, it is not unusual to observe that students are sometimes "harder" on themselves than the teacher is! The goal, of course, is not to see who is "hardest" but for the student to become progressively more effective at honest self-appraisal and productive self-improvement.

Now have a look at the two rectangles below the rubric. The first allows the teacher, a peer, or the student to offer comments, provide feedback, or raise questions. The second box is intended for students to set goals or plan

FIGURE 5.7

Rubric Format for Feedback, Self-Assessment, and Goal Setting

	Title	Labels	Accuracy	Neatness
3	The graph contains a title that clearly tells what the data show.	All parts of the graph (units of measurement, rows, etc.) are correctly labeled.	All data are accurately represented on the path.	The graph is very neat and easy to read.
2	The graph contains a title that suggests what the data show.	Some parts of the graph are inaccurately labeled.	Data representation contains minor errors.	The graph is generally neat and readable.
1	The title does not reflect what the data show OR the title is missing.	The graph is incorrectly labeled OR labels are missing.	The data are inaccurately represented, contain major errors, OR are missing.	The graph is sloppy and difficult to read.

Comments:

Goals/Actions:

actions to improve their future performance based on the feedback from the rubric. Used in this way, the rubric moves from being simply an evaluation tool for "pinning a number" on kids to a practical and robust vehicle for feedback, self-assessment, and goal setting.

Educators who provide regular opportunities for learners to self-assess and reflect often report a change in the culture of the classroom. As one teacher put it, "My students have shifted from asking, 'What did I get?' or 'What are you going to give me?' to becoming increasingly capable of knowing how they are doing and what they need to do to improve."

A Final Thought

Effective assessment practice is a cornerstone of teaching individuals for understanding. Effective assessments serve not only as indicators of student understanding but as data sources enabling teachers to shape their practice in ways that maximize the growth of the varied learners they teach. Effective assessments are not only indicators of student success with content goals but a dynamic part of the instructional process. Furthermore, effective assessment practice not only measures students but assists them in becoming evaluators of their own learning.