From: Educative Assessment, Grant Wiggins, Josey-Bass, 1998

Providing Ongoing Feedback

If assessment is to improve performance, not just audit it, the techniques of measurement must be accompanied by quality feedback provided to learners. The feedback needs to be of two kinds: in addition to better feedback *after* the performance, feedback must also be provided during (concurrent with) the assessment activities, as discussed in the previous chapter. While we know that helpful feedback enables a performer to revise performance to meet high (yet clear and stable) standards, the novelty proposed in this chapter is that we should stop construing feedback and its use as what occurs *between* assessments and start construing it as at the core of what we need to assess. In other words, we must come to see deliberate and effective self-adjustment as a vital educational outcome, hence more central to how and what we test.

Feedback is not merely useful, in other words. It is an essential part of any *completed* learning. As William James described it one hundred years ago, we "receive sensible news of our behavior and its results. We hear the words we have spoken, feel our own blow as we give it, or read in the bystander's eyes the success or failure of our conduct. Now this return wave . . . pertains to the completeness of the whole experience."[1] This emphasis on the "whole experience" is significant because so much of the student's intellectual experience is, unfortunately, fragmentary. Essentially, constant isolated drill work and testing without concurrent feedback means that answers are isolated from actual effects, causes, and purposes in the student's school experience. That is why the "reality therapy" of authentically

contextualized assessment is no luxury. One cannot improve or learn to improve unless one knows how one is doing in performance.

This may seem commonsensical but in traditional school testing it is a radical idea to give the student self-evident feedback while she or he is still being tested or as one stage in a multipart assessment that requires the student to use feedback. Indeed, in many instances it would be considered cheating or perversion of the validity of the results (just as it is also considered radical to give students before-hand knowledge of the tasks and criteria by which their work will be judged). But safeguarding the core premise that assessment should improve performance, not just audit it, requires that assessment embody and demand self-adjustment based on good feedback. To state it more pointedly, we cannot know if students are competent unless and until we see them respond to effects and results.

In Chapter One I asked you to imagine what getting a driver's license, playing in a band competition, and being evaluated as a teacher would be like if these tests were designed to be like most school tests. Here is a similar analogy that suggests just how dysfunctional the present system of testing divorced from feedback is for student learning and improvement: Imagine that the school bas-ketball season consisted of one game, played on the last day of the year, and that the players do not know which of the hundreds of plays they have learned will be tested. Imagine further that this one game is not basketball as the adults play it, not basketball as profes-sional coaches and players see it, but consists instead of a series of drills (a different selection of isolated moves and plays each year) devised by measurement experts and valid to them. Imagine a scor-ing system therefore fully understandable only by the assessors, not by the student players and the coaches. Finally, imagine further that students do not know until weeks later which plays, which have been taken out of game context and isolated from the shooting and making of baskets, have been judged successful. *Who would improve at the game under these conditions? Who would understand the game and his role—the purposes at work—in such an assessment?* Yet this is what typical testing does. The so-called feedback it provides cannot be easily deciphered or used by players and coaches to carry out real performance or real improvement.

To grasp from the learner's point of view how debilitating such a system can be, consider the remarks of recent test takers who expe-rienced the latest innovation in large-scale testing: so-called adap-tive computer testing. It would seem that a test that adapts its questions to your responses in order to better ascertain the level of difficulty and complexity you can reliably achieve would be wel-come. Not so, as far as the test-taker is concerned, especially due to the complete absence of user-directed feedback in such testing:

Only one question appears on screen at a time, and test-takers must answer it before moving on. This means you can't skip around or go back to change an answer. . . .

It is a system that cannot help but fluster test-takers. It is almost impossible to tell how well you are doing, and unpleasant surprises can occur. Edie Sagenkahn, a 30-year-old nurse at Johns Hopkins University Hospital in Baltimore, recalled her panic at being cut off less than a third of the way through 265 questions on a licensing exam last year.

"I was just cruising along, and then the computer just shut off after I had answered the 75th one," she said. "I didn't know if I had passed or failed. I just sat there saying, 'Oh, God, what does this mean?'" Later she found out a friend had plowed through 210 questions, and she felt even worse. Two weeks later she learned that she had passed; the computer shuts off early if the test-taker is performing very well.[2]

This example is just an extreme case of a universal problem. Access to and quality of feedback is often the last thing on the test designers' minds (be they in a testing company or in a classroom). They are looking merely to obtain credible results in an efficient way. Regardless of the purpose (such as licensure, college admissions, accountability, or summative assessment of achievement), almost all tests work the same way: the student must address random questions without knowing how they are doing or without being able to clarify the question and its purpose. What, then, is the unwitting lesson of school testing? Answer, hope, and move on—a view antithetical to quality work. How can the student learn to get better and want to get better if testing does not allow it? Never mind the further absurdity of much test feedback coming after the coursework and school year are over. When all formal assessment is viewed as what one does *after* teaching and learning are over—no matter how performance-based the task—the battle for excellence is lost.

Another reason the computer testing story matters is that most teachers have come to design their tests in accordance with large-scale test formats, which are never designed with anything more than superficial auditing in mind. When students' learning is simply audited, they cannot and do not improve much over time. This is also true, ironically, of the standardized tests being held over teachers' heads. As I shall argue in the next-to-last chapter, educators often misunderstand the feedback from indirect testing; local teaching and testing should not imitate the structure of large-scale testing if the goal is better scores. "Teaching to the test" can lead to worse, not better, student performance on standardized tests—in much the same way that student musicians would worsen over time if all they were taught to worry about were isolated fingering exercises and paper-and-pencil questions about their instruments and music.

What Is Feedback?

We do not get very far in assessment reform without realizing that many educators act as though they do not understand what feedback is. 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 I hear is "Good job!" or an 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 the feedback, the more autonomy the performer develops, and vice versa.[3]

Feedback is information about how a person did in light of what he or she attempted—intent versus effect, actual versus ideal performance. In the more formal language of systems theory, feedback is evidence that confirms or disconfirms the correctness of actions. *Facts* about a performer's performance are fed back to him or her, without the addition (or worse, the substitution) of an adult's view of the *value* of the performance. The audience fidgeted when I spoke; that was not my goal. I sought to keep them on the edge of their seats, and saw that my effect did not match my intent. The pasta meal was too salty (which I did not intend) but was cooked al dente (which I did intend). My clients' body language while I presented my proposal told me it was time for some humor, vivid stories, or helpful examples; I adjusted my approach on the basis of that feedback and they became more engaged.

The best feedback is highly specific; directly revealing or highly descriptive of what actually resulted, clear to the performer, and available or offered in terms of specific targets and standards. Put this way, the challenge is not merely one of having more adults at the ready with fancy verbal or written feedback—a solution that we cannot really afford and would not want to encourage. In light of the previous chapter's discussion of reality therapy in testing, the challenge instead is to ensure that performance tasks place students in the position of having to confront the effects of their actions directly, as suggested by the William James comment quoted earlier in this chapter. In fact, a system of learning and testing that requires a constant reliance on adult interpretation of all student work (instead of making the results and their meaning as self-evident as possible) ensures that learning will be slow and unpredictable.

[handwritten margin note: Real feedback helps you get better!]

[handwritten margin note: Definition]

We thus misunderstand the nature and role of feedback in learning if we think of it as primarily adult commentary. The student needs to see that adult commentary is grounded in what works. In order to see this, the student must gain experience from tasks that permit many "loops" that reveal how his or her adjustments do or do not make a difference in terms of a desired effect. Feedback in healthy systems is thus always timely, continual, and user friendly. Among the most obvious examples of situations in which we seek and use system feedback loops effectively are playing computer games, adjusting the shower water temperature, tasting a meal as we cook, and seeking comments from peer reviewers as we write or perform. Or recall how often a music or athletic coach gave you a steady flow of feedback to tell you how your actions caused this or that result, or did something as simple but helpful as pointing out to you something you did ("There! That time you kept your knee over the ball, and your shot on goal stayed low"). The best feedback is purely *descriptive*, in other words. By contrast, most school test feedback is like the children's game of "hot and cold," where the adult's grades and comments amount to saying "You're getting warmer." Students know that they are close to or far from the goal, but they do not know what or where the goal is; what's worse, they never have sufficient opportunities to play the game long enough and fast enough to ensure success on every major test.

We all need such goal-based perceptions and description of our work, even when the final results are obvious. As the description of what we did is fed back to us (or to a coach who translates it), we come to understand better what we did and did not do, and what we need to do more and less of to cause the desired result next time. We can use that information to close the gap between our intent and the effect we produced.

Consider this feedback from a pitching coach to a major league baseball pitcher who recently began to perform poorly: "Of the 28 pitches he threw," said Red Sox pitching coach Joe Kerrigan, "Heathcliff hit his location only eight times. When I see 8–28, there must be something in his delivery that is keeping him from getting the ball where he wants it. On the videotape, it shows he is opening up his stride by about 4–6 inches. His body direction is actually geared to go into the left-handed batter's box. That's actually taking him away from the plate."[4] No praise, no blame, no vague interpretations—just feedback. Heathcliff Slocumb knew that his effect did not match his intent. Merely knowing that he was not throwing strikes and getting batters out was of limited use as feedback, however. He needed to understand why he was getting that result, just as a student whose writing is vague or whose arguments are fallacious needs to understand exactly what needs to be fixed and how to fix it.

Effective feedback can come from many sources. But as the Kerrigan quote reminds us, even when complex performance is at stake, feedback can come from a camera and replayable tape if we know what we are looking for. (Indeed, we might say that one long-term goal of teaching is for teachers to make themselves obsolete as the only feedback givers.) A videotape cannot only tell a pitcher that he failed to stride and release the ball properly; it can also reveal to students that their speech was mumbled and their words were unfocused, or it can reveal to teachers that far fewer of their students were engaged in discussion than it seemed at the time or that Vance's understanding of the quadratic formula was more solid than it seemed in class. Even in the absence of video, performance will improve if we become more habituated to asking When did it work? Why? When did it not work? Why not? The point in this chapter is that assessment must demand that we ask and answer such questions if performance mastery is our goal.

Our goal in assessment reform is thus not merely to design more engaging and authentic tasks but to build in the kind of frequent feedback and opportunities to use that feedback that are found in all effective performance systems. Road signs do this for drivers. Otherwise we would have to stop and ask for directions or check our maps each time we came to an unfamiliar intersection.[5] Software manufacturers do this for computer users (and minimize costly and time-consuming user telephone calls for assistance) by selling self-correcting systems with on-line tutorials and help files. Assessment should be designed to provide such information to students and teachers. When tasks are designed to maximize the availability of direct evidence and to require self-adjustment, we also come to see that using adult assessors at every step is not necessary (a realization that is an important part of feasibility, as discussed further in the last chapter of the book).

Yet even with better designs, teachers often argue that there is no time to do the kind of assessment proposed here. I fear that these teachers have lost sight of their purposes. The goal is not to teach and hope that it works out; it is to optimize results, just as any coach would. Repeated attempts to teach with good models as concrete examples of goals, with descriptive feedback, and with targeted guidance are what cause excellence. It stands to reason that an unending dose of scattered content, no matter how rigorous, will not by itself produce the educational results demanded. How, then, can there be no time for the giving and using of feedback? We might as well say that there is no time for giving feedback to the trumpet player because there is so much to teach about music and trumpets.

Figure 3.1 sums up what feedback is and what it is not.[6]

[Margin handwritten notes: When did it work? Why? When did it not work? Why not?]

Figure 3.1 What Feedback Is and Is Not

Effective Feedback	*Ineffective Feedback*
Provides confirming (or disconfirming) useful evidence of effect relative to intent, for example, a map and road signs; compares work to anchor papers and rubrics.	Provides nonspecific advice, praise/blame, or exhortations, for example, "Try harder," "Your writing is awful," or "Good job!"; a mere score on the paper.
Compares current performance and trend to successful result (standard), for example, the taste and appearance of the food, not the recipe, guarantee the meal will come out as described; student work is compared against exemplars and criteria.	Naively assumes that *process* (instructions, hard work, and advice) is sufficient to reach goal, for example, planting seeds and diligently watering according to package directions does not ensure a successful garden; students given only directions on how to complete assignment, not guidance on specific standards of final products.
Timely: immediate or performer-friendly in its immediacy, such as feedback from audience and conductor during a recital.	Not timely: too long a delay in usability, or too late to use; feedback on a standardized test provided weeks later, in the summer.
Frequent and ongoing.	Infrequent, given once.
Descriptive language predominates in assessing aspects of performance, for example, you made a left turn onto Main St. instead of a right turn; rubrics describe qualities of performance using concrete indicators and traits unique to each level.	Evaluative or comparative language predominates in assessing performance, for example, you made many correct turns and one incorrect turn, or your navigating is greatly improved and better than that of most of your peers; rubrics basically amount to "excellent," "good," "fair," and "poor," with no insight into the characteristics that lead to such value judgments.
Performer perceives a specific, tangible effect, later symbolized by a score that the performer sees is an apt reflection of the effect, such as the score given by a band judge in competition, based on specific criteria; the grade or score confirms what was apparent to the performer about the quality of the performance after it happened.	No tangible effect or useful result is visible to the performer other than a score, such as a grade at the top of a paper handed back; the evaluation process remains mysterious or arbitrary to the performer, no matter how valid and reliable the test and score are to the expert.
The result sought is derived from true models (exemplars), for example, a first grade evaluation of reading is linked to the capacities of a successful adult reader: the reading rubric is longitudinal and anchored by expert reading behaviors; feedback is given in terms of the goal, such as the specific accomplishments of those who effectively read to learn.	The result sought is derived from a simplistic goal statement, for example, the feedback to first grade readers relates only to checklists: the reading rubric is limited to age-grade skills; there is too much feedback in terms of learning to read, not enough in terms of reading to learn.
Enables performers to improve through self-assessment and self-adjustment.	Keeps the performers constantly dependent on the judge to know how they did.

Assessment as Software

Instead of thinking of assessment as a test (and the student as the test subject), we would be better off thinking of it as activity having the goal of getting students to use our designs to produce high-quality work. Thus we might think of assessment as software that enables students to master complex performance "programs." This view blurs the boundaries between instruction and assessment in a useful way, because it helps us to see that taking stock of progress against standards and using that knowledge to improve performance is both what we want to teach and what we want to assess.

Consider the following example of a system that encourages and assesses student self-correction.[7] In Ralph's welding course he works with thirty students at once and uses a system of effective feedback that requires no adult intervention. The first task is straightforward. Each student must produce a ninety-degree corner weld to industry specifications. The standards for this weld are written out on paper, but the feedback system involves something more. Ralph tells the students that when they believe their weld is up to standard, they should bring it over to a table, pick up a magic marker, and write their name on the weld. By so doing they signify, first, that they understand the standards, and second, that they believe their work is up to standard. There is a catch, however. On the table are welds from previous years, ranging from the excellent to the awful. Students routinely come up to the table thinking they have finished the task and then think twice after inspecting other welds. I watched one boy look around furtively to see if Ralph was watching before he snuck back to his station. This is a feedback system that works, based as all good feedback systems are on activating each person's ability to self-assess and self-adjust when he or she gets clear feedback and sees clear standards.

What a student thinks is a fine weld, essay, or musical performance may be shown to be inadequate once the student views the work through the eye or ear of a judge and in reference to comprehensible standards and criteria. And as long as such feedback is timely, it is welcomed rather than feared because it tells students soon enough (before it is too late to change their work) and unambiguously that their intent is unlikely to be realized unless they make a change.

Feedback Is Not Guidance

If a common misunderstanding about feedback is that it is synonymous with praise or blame, an even more common misunderstanding is the view that feedback is the same as guidance. Feedback and

guidance are quite different; they represent complementary parts of a self-correction system. Feedback tells you what resulted from your action; guidance tells you how to improve the situation.[8] When we teach and comment on performance, we tend to give too much guidance and not enough feedback. In fact, we sometimes jump over feedback and go directly to giving advice—guidance that often does not make much sense unless it is seen as a logical response to feedback.

Here is an illustration of feedback—how it differs from encouragement or criticism, how it is not guidance, and why both feedback *and guidance are essential to eventual mastery and autonomy.* My son Justin was four years and seven months old when he first wrote his name without prompting, like this:

He had just a few weeks earlier discovered his control over a pencil and the differences between the letters that he could produce by writing, so like many kids he practiced writing his own name. We happened to be sitting at the table and he was doing work while I was doing work. He finished and looked at me. Fortunately I had seen what he was up to out of the corner of my eye and had time to consider my response.

"That's really interesting, Justin." (I said this with enthusiasm, but he was puzzled by the word interesting.) "What does it say?"

"It's my name!" he said, somewhat surprised by my reaction.

"Hmmm, show me," I said.

He read off each letter: "J, U, S, T, I, N, spells Justin!"

On another piece of paper I wrote N and H as big block letters and then placed them below where he had written his name. "What's this letter?" I asked, pointing to my N.

"N," he said.

"What's this?" I asked, pointing to the H.

"H," he said, getting a little impatient.

"Hmmm. Well, then," I asked, pointing to his letter N, "What letter is this?"

Long pause. Furrowed brow. "But it is not what I wanted!" he exclaimed, looking at his N.

That's feedback: no praise, no blame; just evidence that asks the learner to consider his intent versus his (unwitting) effect, evidence that talks about what the *product* does or does not do. The feedback is a visible fact, not a judgment. Justin could see for himself, once he was shown a standard, that his N was less well-defined than he had

intended. Note also that it is *his* intent we are talking about, not mine. No adult was laying an agenda on him. Feedback causes learners to attempt to improve because it derives from and then addresses their intent to accomplish some specific thing. Justin intended to form a well-written N. What he wrote looked more like an H. He was shown timely feedback that allowed him to compare a well-written H and N and helped him to discover that he had not achieved his intent. Without this feedback, it might have been much longer before he had seen the need to be more careful in his writing. No overt guidance was provided here, and certainly no disapproval in my words or tone: I did not tell Justin what to do about the feedback, and I made no explicit value judgment. I only provided the feedback, that is, I caused him to attend to the gap between his intent and the actual effect of his efforts.[9]

Moreover, when feedback is specific and relates to a learner's intent, the learner does not feel anger, fear, or loss of self-esteem. Learners of all ages crave such feedback because they are put in a position to recognize what they have done and they are empowered to decide what they want to do about it in light of their own objectives. Students can accept even the most sobering and difficult feedback if it is stripped of all subjective and blaming value judgment and cast in terms of their aim. Why? Because the teacher is taking the time to understand and make careful observations about their work and their intent, perhaps even more than *they* did! (At a recent conference on higher education, a professor remarked to me that her athlete students sought and accepted feedback far better than her nonathlete students, even though the feedback to the athletes often indicated more significant problems of performance. This is presumably due to the athletes' happy experience with direct feedback.)

Most students never get this kind of vital feedback. We too quickly tell them that an answer is correct or incorrect. Or we jump to giving advice. Or we spot a mistake, apply a label to it in the margin, and move on. (I have often told the story of Mike, a student who waited until the end of the year to ask his teacher the meaning of the word she kept writing on his papers—"vagoo" was how he pronounced it.[10]) All these strategies overlook the vital moment of genuine learning when the learner sees his or her work through the assessor's eyes and understands a mistake as a mistake, of a particular kind and cause. The moment when the student understands why some part of his or her work is a mistake is entirely different from the moment when the student perceives that the teacher *does not like* that part of the work. In the second case the teacher understands the mistake but the student does not. If students are to see mistakes as mistakes, we must confront them with clear feedback

and constantly ask them to self-assess, like the students in the welding example.

STOP ———————→ At the elementary level the technique of *miscue analysis* in reading assessment assumes the power and importance of feedback loops. In miscue analysis we make readers' strategies and renderings explicit, helping them to see where they succeeded and where they did not and why, to see where a misreading is plausible and sensible and where not, so that both learner and teacher come to understand reading performance better. But we rarely do such miscue analysis at higher grades in any subject, despite its power for the learner, as shown in the following description of how some mathematics students were taught to self-assess their work through an error analysis after a test: "After we graded their tests, students were asked to evaluate their own performance. . . . Each student was required to submit a written assessment of test performance that contained corrections of all errors and an analysis of test performance. . . . We directed our students to pay particular attention to the *types* of errors they made. . . . They were to attempt to distinguish between conceptual errors and procedural errors."[11]

math example —

The teachers found this process extremely useful: "Student self-assessment and the resulting student-teacher dialogue were invaluable in drawing a clear picture of what students were thinking when errors were made."[12] But they also reported that students found the task demanding. In particular, students "had difficulty weighing the seriousness of an error" and seemed "to have difficulty assigning grades to their work. . . . Many had a tendency to weigh effort heavily regardless of the quality of the product."[13]

Many educators who are still unfamiliar with teaching students techniques of self-assessment imagine that this kind of work requires age and intellectual maturity. Not so. Figures 3.2 and 3.3 present scoring rubrics for teacher assessment and student self-assessment that are used in two different elementary schools. Figure 3.2 contains a chart from which the teacher reads descriptions of student performance out loud, after which students place cardboard smiley faces in the appropriate column.[14] (In a videotape on assessment reform, a little girl who was asked why she judged herself to be only "proficient" and not "advanced" in "Response to Literature through Conversations/Conferencing" said matter of factly that "there are some things there that I don't do yet. I *try,* though."[15]) In Fairfax County, Virginia, teachers and supervisors have developed a comprehensive handbook for assessing literacy development. In addition to a narrative section on assessment policies and practices, the handbook contains numerous tools and strategies for assessing student performance, including student self-assessment forms such as the one contained in Figure 3.3.[16] And Figure 3.4 contains an