

9

ASSESSMENT AS DIRECT FEEDBACK TO THE LEARNER

To this point we have been looking at assessment primarily as a means of informing or enlightening the educator: to aid the decision-making process by providing information on how various educational programs, practices, and policies affect talent development. In this chapter we shall examine the ways in which assessment can be used to influence talent development directly by serving both as feedback and as an incentive to the learner. We shall consider both students and faculty members as potential learners.

From the perspective of the input-environment-outcome (I-E-O) model, direct feedback constitutes one of the *environmental variables* that can be used to improve performance. The learner's performance which initially generates that feedback can be viewed as an *input* or *pretest variable*, while the learner's subsequent performance (following the feedback) can be viewed as an *outcome* or *posttest variable*.

Although measurement specialists have never developed formal theories to explain how or under what conditions assessment affects the learning process, there seem to be several implicit theories that underlie much of the discussion and debate about assessment. Since each theory implies a somewhat different approach to assessment, it is important to understand the reasoning behind each one. For simplicity I have identified two broad categories of theory: incentive theories and information theories.

ASSESSMENT AS INCENTIVE: THE CARROT AND THE STICK

Theories based on "carrot and stick" reasoning see assessment primarily as an external incentive to learn that operates both as a reward and a punishment. When it comes to facilitating learning, most of our traditional assessment in higher education can best be justified on the basis of its incentive value. Surely the most common example of assessment based on this particular theory is the grade. Grading in higher education

is involved at several different levels: we grade course assignments, exams, and overall course performance (the final grade), and we aggregate course final grades into a cumulative grade-point average (GPA). I have already argued that grading in higher education is often justified on the basis of its usefulness in screening and certifying, but to the extent that it is regarded as being useful in the learning process, it is supposed to operate primarily as an incentive: students are expected to work harder because they know their performance will be graded. Grades, in turn, are important to the student not only because they can influence subsequent employment and educational opportunities but also because of their implications for the student's self-esteem. Getting good grades is presumably good for one's sense of self-worth, and getting poor grades is presumably detrimental to one's sense of self-worth. Thus grades can operate as both a carrot and a stick, depending on whether the student is oriented toward attaining success or avoiding failure.

While standardized tests and other types of student screening devices are viewed as aids to the talent development process even less often than grading is, to the extent that such tests in their present form might affect learning, they would operate once again under the carrot-and-stick principle. High school students are thus expected to study hard for the SAT or ACT because they want to gain admission to the best (i.e., most prestigious) college. Undergraduates will study hard for professional licensing exams or to prepare for the GRE, LSAT, MCAT, or GMAT because they want to be admitted to the best possible graduate or professional school. Finally, professional school students will study hard so they can pass the state bar or medical licensing exams. In all such situations these examinations can be regarded as external incentives or motivators.

Most assessments of college and university faculty follow the pattern for student assessments: to the extent that they are assumed to have any value in developing faculty talent, they are supposed to operate primarily as incentives. Professors' knowledge that their scholarly work will be assessed by their colleagues presumably serves to motivate them to do more and better writing and publishing, and their awareness that their students will evaluate their course is expected to motivate them to do a better job of teaching. Again, both the carrot and the stick are involved: college professors want to be respected by their colleagues and admired by their students, and they want to avoid losing their jobs because they fail to make tenure or, if they are already tenured, being regarded as "dead wood."

In a similar vein, the reasoning behind many of the state initiatives to mandate assessment activities in higher education (see chapter 11) is based on carrot-and-stick reasoning. To some state officials, student outcome assessment represents a kind of club that can be used to motivate faculty to be "more responsible" for their students' learning

and to compel institutions to be more "accountable" in their teaching activities. Such officials thus believe that institutions and their professors can be forced to do a more effective job of teaching if the results of their pedagogical efforts are exposed to the public. (Assessment programs in Florida and Georgia appear to be based at least in part on such reasoning.) The not-so-subtle implication here, of course, is that current efforts at teaching are *not* very successful. Otherwise, why are the outcome assessments needed?

In other states, mandated assessment activities are viewed more like a carrot (Tennessee being one example). When assessment results show that an institution is doing a good job, it gets extra funding from the state. Of course, once such a program becomes established and most institutions are receiving incentive funding (the current situation in Tennessee), the distinction between carrot and stick becomes increasingly difficult to make. *Not* receiving full incentive funding is thus viewed as a punishment, i.e., a virtual loss of funding. (The pros and cons of state incentive funding programs are discussed in detail in chapter 11.)

ASSESSMENT AS INFORMATION: THE FEEDBACK PRINCIPLE

A very different way of looking at assessment is to regard it as a way of generating information that can facilitate the learning process. This approach is based on a well-established principle of learning called feedback or knowledge of results. Many hundreds of studies of human learning have shown that the learning process can be substantially enhanced if the learners have appropriate knowledge of results showing how much progress they have made and pointing out specific areas where additional work is needed. Without such feedback, learning can be very slow and difficult or, under some conditions, virtually impossible.

To illustrate how the feedback principle works in practice, let us consider a common learning situation in higher education, the freshman course in English composition, and contrast the feedback approach with the incentive approach to assessment. Assume that we have two students of equal writing ability but whose first essays are graded by different teaching assistants. Student A's essay comes back with just a grade, C-, and perhaps some "incentive" comments such as "You need to work harder in this course," "This is a very weak essay," "Not bad for a first try, but you need to improve." In contrast, Student B's essay comes back with no grade but with a series of specific comments: "This first paragraph should tell the reader briefly what the whole essay is about," "This is not a complete sentence. Rewrite it to make sure it has a subject and a verb," "Try writing shorter sentences," "This is a very interesting idea; try explaining it a bit more." Whereas the information given to student A might serve as an incentive to work harder, student

B has a much clearer idea of what he should do to improve his writing. There is no reason, of course, why both the incentive and the enlightenment principles cannot operate simultaneously; the point to keep in mind is that the two principles involve different mechanisms (reward and punishment versus knowledge of results) and imply different kinds of assessments (evaluative judgments versus feedback to steer the student's future efforts in particular directions).

Perhaps the best examples of the feedback principle as applied to the assessment of student learning are the specific narrative comments provided to students on exams, essays, term papers, and other student work. Some innovative institutions—Hampshire College and the University of California at Santa Cruz, for example—actually use narrative evaluations in place of the traditional course grades and GPA. Instead of a decimal number, students receive a written evaluation that attempts to touch not only on the overall quality of their work but also on specific strengths and weaknesses. Such comments usually cover cognitive skills (thinking, logic, writing, speaking, analysis) as well as work habits and motivation. Moreover, where appropriate, these evaluations also include specific suggestions for improvement.

Assessments of faculty performance are seldom done to provide "knowledge of results," with one notable exception: the pre-tenure review. In some institutions the work of assistant professors is reviewed in detail after their first three or four years of employment. One important purpose of such reviews is to identify any significant deficiencies in the assistant professor's performance far enough in advance to give the candidate an opportunity to remedy them. However, many academics seem to believe that if such reviews are even mildly negative (especially with respect to the candidate's scholarly work or research), the candidate's prospects for ever making tenure are bleak indeed. Negative reviews are thus taken to mean that the candidate should begin to seek employment elsewhere. Such beliefs, unfortunately, tend to undermine the potential educational value of this feedback.

One important form of assessment in higher education that is done *primarily* to provide feedback is the institutional accreditation process. Except for newly established institutions that are seeking accreditation for the first time, the accreditation process is seldom concerned with the question of whether the institution being assessed should be re-accredited. Consequently, a positive decision on reaccreditation is a foregone conclusion in the vast majority of cases. What is really involved in most such visits is a thorough review of the institution's curriculum, facilities, programs, policies, fiscal condition, and governance by an expert team of colleagues from peer institutions. The basic aim of most accrediting visits is thus to "enlighten" the institution concerning its major strengths and weaknesses. Most accreditation teams make a number of specific recommendations for how the institution can strengthen its programs and operations.

INCENTIVES OR FEEDBACK?

While there is no reason that a given assessment activity cannot simultaneously utilize both the incentive and feedback principles, it is important to realize that these two principles are based on quite different conceptions of the assessor. Thus, advocates of the incentive approach believe that external rewards and punishments are needed to motivate students and faculty to develop their talents. The feedback theory, on the other hand, implicitly assumes that students and faculty naturally want to learn and to develop their talents, and that what they need in order to do this is good information about their progress, their specific strengths and weaknesses, and the specific types of activities that are most likely to contribute to their future development.

So far we have been focusing on theories of how assessment can facilitate the talent development process directly. We have noted a few instances in which assessment in higher education is thought to operate on the feedback principle, but the educational rationale underlying most of our assessment activities, especially testing and grading of students and reviews of faculty performance, seem to be based more on the rewards and punishments of the incentive principle. I have already noted that incentive assessment can also have feedback value, but there are at least two instances in higher education in which the power of assessment to enlighten is seriously compromised by its use as an incentive.

The first of these is the course grade. The fact that professors must grade their students as well as teach them makes it difficult for students to develop the trust that is sometimes necessary for them to ask for the kind of feedback they really need. The incentive principle encourages students to impress their professors with their knowledge and competence so as to get a good grade. At the same time, it discourages them from exposing their ignorance, so to speak. If students are confused or uncertain about some aspect of the course, they may be reluctant to seek appropriate feedback because they are afraid that the professor will evaluate them negatively. For the past fifteen years I have taught a very technical and difficult graduate course in statistics and research methodology, and the biggest pedagogical problem I have is convincing students that they will not be punished or judged negatively if they are completely honest with me about what they do not fully understand. Unfortunately, our predilection for incentive-oriented assessment in education has conditioned many of our students to view the professor or teacher more as someone to impress or manipulate than as someone who can help them learn.

The other assessment activity in which the potential educational value of feedback is compromised by the incentive principle is student evaluations of teaching. Many of us who teach are in a position to benefit substantially (develop our teaching talents) from end-of-course eval-

uations by students. However, because the results of such evaluations are usually made available to others and used in personnel decisions, we often focus more on getting good ratings than on actually learning from the ratings (Gleason, 1986). (This situation is directly analogous to that in which students are motivated to study for the test rather than to learn the material.) One specific danger in such ratings is that we teachers will be tempted to "go easy" on our students for fear of getting poor ratings. Thus, we might give fewer assignments, easier exams, or fewer exams in order to get more favorable ratings. Or, we might be tempted to do more entertaining and less teaching in the class. The problem, in short, is that when such ratings are used as incentives, we are being encouraged to manipulate them rather than learn from them.

EFFECTIVE ASSESSMENT IN THE CLASSROOM:
A PROTOTYPIC EXAMPLE

Since the classroom provides our greatest opportunity to influence student learning directly, let us now consider how classroom assessment and feedback can be used to enhance the learning process. Ironically, most principles of good assessment for learning can be illustrated by looking at some of our "softest" fields, especially the fine and performing arts. Let us take a prototypic example from the arts: the piano lesson. By examining in some detail just what happens during a typical piano lesson, we can discover virtually every principle by which assessment and feedback can be used to enhance the talent development process directly. As we discuss the lesson, I will spell out each principle in italics.

To begin with, both the teacher and the student are interested in improving the quality of the student's piano playing. In other words, *both the teacher and the student are committed to a common goal or set of goals*. These goals, at least in part, have to do with talent development: growth or improvement on the part of the student. Most students, however, are also interested in another type of goal: *that the learning process itself should be enjoyable, gratifying, meaningful, interesting, or in some other respect pleasurable or rewarding*. Piano teachers sometimes forget this, and when they do, learning the piano from the perspective of the student often becomes boring, unpleasant, or pure drudgery. When all the fun and enjoyment is taken out of the learning process, the only remaining incentive is the learning outcome itself (improvement in playing ability), and often this rather remote goal is simply not enough to sustain the student's interest in taking lessons and practicing. Again, such problems underscore the importance of having the teacher and the student committed to a common set of goals.

What about the learning process itself? Let's first look at what the teacher does. Once the teacher has defined the task ("practice this scale," "learn that piece," "try playing it like this"), the student is asked to perform, and the teacher *watches and listens*. Basically, what the teacher is doing here is *assessing the student's performance*. Clearly, *the performance assessments should be relevant to the shared goals of the learning process*. While it is possible to make a single overall judgment of the student's performance (excellent, good, fair, poor), performance assessments are inevitably based on several criteria rather than one. Thus, the student's playing can be judged in terms of accuracy (playing the right notes), tempo, dynamics (loudness and softness), touch, interpretation, and so on. Such *outcome* assessment might also be coupled with process assessments, whereby the teacher sees how the student sits or holds her hands, observes her emotional state, or inquires about how she practices between lessons. In short, *the teacher's assessments of student performance should be multidimensional and can involve observations of the learning process (environment) as well as of learning outcomes*.

The information obtained through the teacher's assessments is then used by the teacher to generate *feedback* to the student, usually in the form of spoken comments and suggestions. Sometimes this feedback consists simply of the raw assessment data ("you should have played an A-flat here instead of an A"), sometimes it consists of direct suggestions for improvement ("try counting out loud while you play"), and sometimes it is intended merely to serve as encouragement ("you played it much better that time"). Whatever the form, *assessment and feedback is intended to serve the goals of teaching and learning: to facilitate student learning and talent development and to make the learning process itself more rewarding*.

Once feedback has been provided, the student typically is asked to perform once again, and the assessment process continues. When feedback from the teacher does not serve its intended purposes (performance does not improve, the student becomes frustrated or discouraged), the teacher may simply repeat the feedback ("you still played an A instead of an A-flat; try it again"), or she may try using a different form of feedback ("try playing this passage more slowly"). In other words, *assessment and feedback should be an ongoing, iterative proceeding that is integral to the learning process rather than a one-time activity carried out only at the end of the learning process*. It would appear that, in the fine and performing arts at least, much of what we refer to as "teaching" activity consists of providing assessment and feedback to the student. Whether assessment and feedback should have the same priority in teaching other academic fields is perhaps debatable, but there seems to be little question that *assessment of student performance and providing feedback to students should constitute a significant part of any teacher's pedagogical activities*.

Let's now look at the piano lesson from the student's perspective. The nature of almost any performing art is that the performance itself generates direct feedback to the student. The student, in other words, can listen to what she plays (feedback) and make a judgment about it (assessment). It is thus possible for the student to learn to play the piano without the aid of a teacher (and, in fact, many people do), since a good deal of feedback is automatically built into the performance process. In short, *students can learn by generating their own feedback and assessment*. However, if the student is ever going to achieve a high level of proficiency in playing, a substantial amount of assessment and feedback from others (fellow musicians, friends, teachers) is almost always required. Most "learners," in other words, eventually need "teachers" to provide assessment and feedback.

It is difficult to overestimate the importance of assessment and feedback in the learning process. Consider what would happen if the feedback were eliminated. Imagine what it would be like for a deaf person to try to teach someone else how to play the piano. Or imagine how difficult it would be for a deaf person to teach herself how to play. Similar problems arise in other fields of the arts. How could a blind person learn how to paint or to teach anyone else to paint or draw? And how could you teach somebody to write if you could never read their writing?

It might be added here that the importance of assessment and feedback to the learning process is just as great when we reverse the roles and view the teacher as a learner. How do people learn to be good teachers? Again, we see the essential role of assessment and feedback. The piano teacher can make assessments of the student's performance, provide feedback, and then observe the student's performance once more. If the performance gets worse, fails to improve, or improves too slowly, the teacher can try something else. Gradually, the teacher learns what works and what doesn't work, and under what conditions. Note, however, that the student's performance in this instance serves as feedback to the teacher. The teacher still does the assessment, but without such feedback provided on a regular basis, it is difficult to see how anyone could learn to be an effective piano teacher.

In many respects the need for assessment and feedback in learning any performing art is self-evident. Indeed, if the performing arts constituted the whole of our curriculum in higher education, there would be little need for a book such as this one, since the role of assessment and feedback is fundamental to the implicit theories that govern the activities of all teachers of the performing arts. No performing arts teacher in her right mind would employ techniques that contradict or ignore these principles. However, the central role of assessment and feedback is often overlooked in the implicit theories used by teachers in most other fields. If professors in these other fields were to apply the principles of

effective assessment and feedback more fully in their own teaching activities, the educational impact of our higher education programs would improve substantially.

I have chosen the performing arts to illustrate the importance of assessment and feedback in the learning process for two reasons. First, the essential role of assessment and feedback is patently obvious in the arts; it would be next to impossible for either teachers or students to learn without it. Second, in the performing arts it is easier to see the importance of timing in providing feedback. Imagine how difficult it would be for the beginning piano student to benefit from feedback that came two weeks after she performed. But problems in timing can also occur when there is too much feedback provided too soon. For example, if the teacher stops the student from playing when there is the slightest error in any aspect of the performance, the student can quickly become frustrated and discouraged. Or, if the teacher provides too much feedback at one time or asks the student to attend to too many different things while performing, the student can become confused and not really be able to benefit from the feedback.

FEEDBACK IN OTHER FIELDS

What is the nature of feedback in the humanities, social sciences, and natural sciences? Unfortunately for the learner, most teachers in these other fields have chosen to employ teaching methods that do not, in themselves, generate timely and appropriate assessment and feedback for either the student or the teacher. This is partly due to the nature of the fields themselves, and partly an economic matter—large lecture sessions do not readily lend themselves to individualized feedback.

What can professors do to simulate the kinds of feedback available to those who teach in the performing arts? A unique approach to assessment in the classroom is what Patricia Cross (1989) has called "classroom research." Cross's basic idea is for classroom teachers to develop a series of small-scale assessments that can be given regularly in a class to provide information about what and how students are actually learning. Cross and Angelo (1988) have developed a detailed handbook of such classroom assessment techniques for use by teaching faculty. Three different kinds of assessment techniques are presented and described: those for assessing intellectual development, those for assessing students' self-awareness as learners, and those for assessing student reactions to both the teacher and the course.

Part of the difficulty with academic fields outside the arts is that the "talents" that are being developed are not as salient as performance talent in music, art, or theater. One cannot directly observe the student's critical thinking ability, cultural understanding, esthetic appreciation, or

knowledge of history or chemistry. For this reason, student assessment in these fields has traditionally come to rely on indirect measures called tests. What is especially significant about testing in higher education is that most of it is *not* done to provide students and faculty with feedback to enhance the talent development process. Rather, as we have already seen, testing is done primarily to produce course grades and to sort, select, classify, and certify students. As a consequence, both the nature of most tests and the timing of their administration are ill suited to the teaching-learning process and therefore of limited value in furthering the institution's talent development mission.

To illustrate these last points, let's look at how we test students in the light of the (italicized) principles set forth in the preceding paragraphs. For purposes of this discussion we will consider the kinds of assessment and feedback that the student is likely to encounter in a typical lower-division undergraduate course offered in a large public institution (where the majority of undergraduates are to be found). While the problem illustrated here may not occur as frequently in smaller institutions or in upper-division courses, there are few courses outside the performing arts where some improvement could not be achieved through more diligent application of the principles of effective assessment and feedback.

SHARED GOALS: TALENT DEVELOPMENT AND SATISFACTION. Are the teacher and student in the typical psychology or history course committed to a common set of goals? Not necessarily. The student may merely want to be entertained or to satisfy a course requirement while the professor may be more focused on helping students learn. Or, the professor may be more interested in impressing students with his knowledge whereas the students may be focused more on getting a good grade.

RELEVANCE OF ASSESSMENT AND FEEDBACK TO SHARED GOALS. Even if professor and student were committed to the common goal of talent development, the single most common assessment procedure, the course final examination, would be of limited value because it does not really say much about what has been learned (i.e., about talent development). This and the preceding principle are, of course, closely linked: without a common student-faculty commitment to talent development, it is highly unlikely that assessment and feedback will be used to further this purpose.

MULTIDIMENSIONALITY. Students in many college courses receive little feedback beyond a single score on a final examination (the number or percent of correct answers) or worse yet, a relativistic assessment in the form of a letter grade (A-, C+, etc.). Even when the student prepares a term paper rather than (or in addition to) an exam, the feedback often consists of little more than a final grade or a few written comments.

INCLUSION OF "PROCESS" DATA. This category includes information such as how the student behaves in class (interaction with others, note taking) as well as out-of-class behavior such as reading and study habits and use of the library. Only in very rare instances do faculty members even bother to gather such information, much less use it in providing feedback to students. It is true that many institutions now have programs or centers to enhance students' learning skills, but such activities are usually not directly linked to courses, and student participation is usually voluntary and therefore infrequent.

ENHANCEMENT OF LEARNING. Perhaps the best assessment for learning that goes on outside the performing arts occurs in laboratory work or with the homework assignments that are common in the natural sciences. Laboratory work is usually monitored closely, and homework (e.g., problem sets) is usually evaluated on an individual basis. Otherwise, about the only widely used assessment approach that might be useful in enhancing the learning process is the mid-term exam. Depending on the form of such exams and the type of feedback provided, such assessments have the potential for being of significant value in the learning process. The typical final exam, of course, comes far too late in the process.

ASSESSMENT AS AN ONGOING, ITERATIVE PROCESS. Except for mid-term exams and lab work and homework assignments in the natural sciences (above), most assessment and feedback is a one-shot activity conducted in connection with specific courses. Daily "pop" quizzes are perhaps the only exception, and they are seldom used. Furthermore, assessments from different courses are seldom compared and contrasted over time. Thus, for any student it is very difficult to determine how much or what type of talent development is actually occurring.

FACULTY TIME AND EFFORT DEVOTED TO ASSESSMENT AND FEEDBACK. Outside the arts, the greatest degree of faculty involvement in assessment and feedback probably occurs in courses in English composition. Among professors who teach most large undergraduate lecture courses, however, only a very small fraction of total teaching time is devoted to assessment and feedback. Indeed, in many large undergraduate courses, professors devote no time to assessment and feedback, relying entirely on teaching assistants to administer examinations and grade papers. When professors do conduct their own assessment activities, they often utilize multiple-choice tests that can be scored and graded by machines.

STUDENT-GENERATED FEEDBACK. The most common form of student-generated feedback occurs with term papers and other written assignments: students can read what they have written and adjust their writing accordingly. Students can also generate feedback from each other

by sharing and critiquing written assignments. Some institutions (e.g., Evergreen State College in Washington) actually structure their general education courses in this fashion. But formal opportunities for students to generate assessment and feedback from professors are relatively rare, especially in large undergraduate lecture courses. Nevertheless, an enterprising student in almost any type of institution can generate a considerable amount of feedback by staying after class to talk to professors and by meeting with professors during office hours. Unfortunately, few students take advantage of such opportunities, nor do most institutions encourage students to do so.

OPTIMAL TIMING OF FEEDBACK. Most feedback occurs at the worst possible time: at the end of the course. Once the student finishes the final paper or final exam, interest in any meaningful feedback (beyond the final grade) diminishes considerably, as the student goes on vacation, prepares for the next term, or graduates. As already mentioned, mid-term exams are well timed, but what may be the best time for assessment—at the beginning of a course—is seldom used.

In summary, this overview indicates that the best principles of assessment and feedback are seldom followed or applied in the typical lower-division undergraduate course. And if we include standardized tests in our review, the picture is even more discouraging. The most widely used tests, of course, are the national undergraduate admission tests (the SAT and ACT) and the graduate and professional school admissions tests (GRE, MCAT, GMAT, LSAT). The feedback from these tests consists primarily of norm-referenced scores that tell the student how well he is performing only in relation to other students. The feedback is multidimensional (in that most tests generate more than one score), but the student is generally at a loss to explain how or why he performed as he did on the various subtests. This is especially true for "aptitude" tests like the SAT and GRE. Finally, the normative feedback tells the student almost nothing about what or how much has been learned over time (no two tests produce scores that can be compared to derive change measures), and even less about what the student might do in the future to improve his performance.

Feedback from lower-division assessment activities is of equally limited value to the professor. Standardized, norm-referenced test results tell the professors virtually nothing about the effectiveness of their pedagogical efforts, and course grades are not much better. Final examination results offer somewhat better information, especially if they are aggregated across students in some fashion to provide a picture of the class's specific strengths and weaknesses (a practice that is, unfortunately, greatly underutilized). Final exam results would, of course, be much more informative if they could be compared with pretest results given at the beginning of the class. Perhaps the best single source of feedback for the professor is the student course evaluations that are now widely used in many institutions.