Assessment

What's this doing here?

I'm sure that many readers will find this essay somewhat out of place, appearing near the front of this collection. Most books about teaching deal with issues of assessment at or near the end, an ordinal position that mirrors the perceived chronology of teaching in which assessment and evaluation come into play only after instruction is over. Plan, teach, evaluate, right?

Let's take a moment to rethink this idea. Although it's certainly true that assessment of student learning is often considered a culminating activity, thinking about assessment only in this narrow way is decidedly unproductive.

Many teachers begin teaching without asking and answering the important question of how they'll know that students have learned what they think they've taught. The issue of assessment to some may appear so obvious as to be implicit in any instructional planning, but the extent to which teachers and students continue to be surprised by students' performance on tests, auditions, or juries, or by students' inability to demonstrate what they know in contexts other than those presented to them directly, is ample evidence of the fact that questions of assessment are not tacitly understood. They're tacit all right, but they're not very well understood.

Assessment is inextricably related to the goals of instruction, so the time to begin thinking about assessment is in the planning stages, before instruction actually begins. Implicit in every goal statement is the question of how students will demonstrate that they've accomplished the goal. This is no small point, and it brings into sharp focus the need for

precision in our thinking about planning and teaching. How will students demonstrate that they've accomplished the goals we set for them? If we want students to understand key signatures in common practice tertian harmony, for example, how will they show us that they understand?

The current controversies over standardized testing are emblematic of this fundamental question: What should we accept as evidence that students have learned? Ancillary questions that appear often in discussions of public policy are intimately related to this fundamental issue. What should be the criteria for determining the success of a school? How do we know that our education dollars are well spent?

Decompartmentalizing teaching and learning

Most conceptions of education partition teaching practice into curriculum, instruction, and assessment. Teachers are supposed to decide upon their goals, plan how to attain them, and then evaluate whether students learn what's taught. This point of view seems logical enough, until you begin to think more deeply about what all of this means in precise terms.

Even the administrative structures of colleges of education reflect this compartmentalized way of thinking. There are departments of curriculum and instruction that deal with the organization and delivery of subject matter content. This is where the pedagogues live. The psychometricians—the test people—live elsewhere, usually in departments of educational psychology. And although most students of education must take coursework in curriculum, instruction, and assessment, these topics are usually addressed separately by specialists who focus on each apart from the others. Too bad, that. Too bad because what to teach (curriculum), how to teach (instruction), and determining whether students have learned (assessment and evaluation) are not separate at all, but are inextricably interwoven.

Now, many of my colleagues in education will accuse me of creating a straw-man argument, because no one teaching a course in curriculum, instruction, or assessment and evaluation believes that these topics are entirely separate from one another. Of course how one teaches is related to what one

teaches, and how and what prompt the question of how well. My concern is that, although acknowledgement of the interrelatedness of ideas may exist in the minds of my colleagues, it is clearly not in the minds of most students, nor does it seem to be a prominent feature in the thinking of many practicing teachers.

It's important at this point to separate assessment, which is a process of gathering information, from feedback and grading, which are forms of conveying the results of assessments to others. This is no small point. Assessment and grading are not synonymous, even though many students, teachers, and principals consider them so. Assessment is "finding out." Feedback and grading are "communicating what you've found out."

Many folks in education go to great lengths to explain how the terms assessment, measurement, and evaluation are different from one another. While it's true that these terms can express subtly different ideas, I find the usual time spent explaining the differences rather pedantic and generally unhelpful. Put most simply: *assessment* is the measurement of a learner's performance; *evaluation* describes the learner's performance in relation to other learners or according to some continuum of graduated labels. "You made 4 out of 28 free-throws (assessment), which stinks (evaluation)."

In this essay I discuss assessment more broadly as a process of data collection—a process of finding out what students know and what they've learned to do. The importance of this perspective is that it emphasizes the finding-out aspect of assessment and illuminates the usefulness of assessment at all points in the instructional process. Assessment is not merely a culminating activity. Assessment is an ongoing activity, one that should be at the fore in a teacher's thinking from the first moments of goal setting and throughout the process of planning and implementing instruction.

The grading issue is another can of worms entirely. Unfortunately, the fact that assessment is often conflated with evaluation and grading leads many teachers to dwell on devising unnecessarily elaborate mechanisms for awarding and

revoking points and privileges, computing averages, and converting number scales to letter grades, a process whose complexity convinces other teachers to simply avoid the whole mess for as long as possible. Students, teachers, administrators, and parents who express feelings of anxiety, fear, aversion, frustration, indignation, disgust, or downright hatred toward assessment of learning are for the most part concerned not with the assessments themselves but with the *consequences* of assessment. It's the grades, the test scores, and the audition results, and what those evaluations mean for the future that gets everyone exercised about assessments in school.

This is not an intractable problem, but its solution begins with separating the finding-out aspect of assessment from the consequences of conveying what's been found out in the form of grades, chair placements, admissions decisions, and scholarships. This is no small point, and I invite you to think carefully about your own feelings about assessment, evaluation, and grading from both sides of the grade book, as student and as teacher, and to try to understand the proximate and ultimate causes of your enthusiasm or distress. I too find the grading issue problematic, so I'll avoid the mess for now and will return to the evaluation issues later.

Collecting data

Assessment is data collection, and all of us, whether or not we're teaching, make assessments about what's going on in our environments all the time. We observe the physical features of the spaces in which we move, the sounds that are present, the physical states and reactions of our own bodies, the content of our private thoughts and feelings, and the behavior of others who share our space, all of which we assimilate, synthesize, and interpret to create what seems to us a coherent view of the world in which we find ourselves. Of course, all of this happens without much conscious effort on our parts. Our minds are built to do just this, to receive and interpret the signals in our environment so as to help us formulate decisions about what to do next.

No two people have exactly the same interpretation of a given experience because each of us brings to each new

encounter his own history of past experiences, and this history inevitably colors what we see, hear, and feel in the present. Interpretations differ between individuals also because each of us is selective in his perceptions; that is, we tend not to all of that which exists in our perceptual fields. There is simply too much to take in at once, and, as a means of accommodating the flood of information that is available to us at any moment, our minds select—sometimes consciously, often not—what seems to be most important and advantageous to attend to. This perceptual selectivity is the combined result of some hard wiring in the human perceptual apparatus and our learning from past experiences, a combination that creates expectations and heightened sensitivities to the world around us.

The combined effects of the hard wiring and experiences are readily exemplified in the development of human language skills. The research data illustrating the perceptual acuity of human infants for speech sounds (and music) are stunning, revealing that infants possess keen sensitivities to sound and are capable of auditory discriminations that most adults would not imagine possible. All healthy infants begin life with the ability to perceive all of the sonic variations in human speech, for example. As infants live and grow in an environment filled with language, their perceptual sensitivities are modified by their experiences, including the relative frequencies of the speech sounds they hear. Human infants learn where words begin and end in the seemingly uninterrupted auditory stream that is human speech, a remarkable feat which they accomplish through comparisons of statistical probabilities of the occurrences of phonemes. Through even the first months of life, infants learn which speech sounds convey information and which do not, and as a consequence they learn to pay attention to the sounds that matter and to ignore the ones that don't. As infants experience the sounds of speech, they learn systematically to attend to sounds that are necessary for understanding, like the difference between the e sounds in bet and beat and to ignore acoustic variations that have no meaning, like the differences in pronunciation among individual speakers.

This learning results in physical changes in the brain that become more or less permanent after the first few years of life.

Once the important discriminations are learned, it becomes almost impossible for the child to hear the unimportant variations because the brain has lost the capacity to do so. Thus, the reasons that most non-native speakers cannot learn to speak a language accent-free after the age of six or seven involve not only the speech apparatus but the auditory discrimination apparatus as well. Japanese speakers, for example, find it next to impossible to pronounce r's and l's, and often confuse the two sounds in English. Because the r's and l's of English are not present in Japanese language, but are variants of a single phonemic category, the brain's capacity to make the necessary discriminations between these sounds is lost early in a Japanese speaker's development. It's not only that the lips and tongues of Japanese cannot do the tricks that produce r's and l's, the problem stems from the fact that native Japanese speakers cannot *hear* the difference between r's and l's.

All of this is to illustrate the fact that we are all active data [illegible] life experience beginning in infancy, and the data we receive and interpret influences how we think, what we do, and how we feel. This may seem like a trivially obvious observation at first, and only peripherally related to the topic of this essay, but it has profound implications for teaching and learning, since we and our students are collecting and interpreting data all the time. Of course, most of this data collection is informal and unstructured, but it is of consequence nevertheless, because the data that we take in potentially affect what we do, in subtle and not-so-subtle ways.

Assessment in teaching

As we move through each lesson, rehearsal, and class, we make innumerable observations about our students' knowledge, skills, and attitudes. This informal, ongoing assessment guides our decision-making moment to moment as it illuminates the extent to which our students (1) understand what we're talking about, (2) can do what we ask of them, and (3) are interested enough to care one way or the other.

The extent to which we as teachers are consciously aware of the information we assimilate and the extent to which it influences our subsequent behavior varies among individuals

and among circumstances. Some teachers are very perceptive about the goings-on in their classrooms and use the information they gather from students to skillfully construct sequences of proximal performance goals, head off problems, elicit student effort and attention, and increase the likelihood that students will successfully accomplish what the teachers have in mind. Others, who fail to recognize the wealth of information available to them, or who fail to interpret the information accurately, are simply unable to explain why students just can't understand what the teacher's trying to put across, or why the students can't do what the teacher asks, *or why they can't just show a little interest! Good grief!*

Of course, a limited ability to size up a situation in which a teacher finds herself is an insurmountable impediment to teaching well, because at the very heart of skillful instruction is the keen perception of what's going on in the room. The education jargon has appropriated the term "with-it-ness" (no, really) to describe this knowing aspect of teaching. Teachers who are said to be with-it know what's going on around them, accurately perceive students' knowledge, skills, and attitudes moment to moment, and incorporate this information into the ongoing process of instruction. Many people engage the same type of skill in informal social settings, company planning meetings, and financial negotiations. These are the folks who are said to be "good with people." Educators and other scholars who buy into Howard Gardner's notion of "multiple intelligences" would say that these people possess a high degree of "interpersonal intelligence." Whatever. The point is that these types of assessment skills are central to successfully leading others to prescribed goals.

The absence of with-it-ness renders effective teaching an impossibility for even the most knowledgeable practitioner of a discipline. The advent of so-called teaching machines in the 1950's was met with optimistic (?) predictions that teachers would soon become less central to the teaching-learning process, if not superfluous, and that students could simply learn from well-programmed instruction absent the intervention of a live teacher. Although the burgeoning sophistication of computer technology has moved the original notion of teaching machines well beyond what anyone could

have imagined in 1956, there remains a need for human beings to deliver and guide instruction. Education at the dawn of the 21st century is flush with a dazzling assortment of technological advances, yet we still need human teachers.

What teaching machines then and now lack is the capacity to accurately assess students' knowledge, skills, and attitudes, and to incorporate this information into the ongoing process of instruction *in the moment*, modifying instructional tasks in ways that lead learners to successful conclusions. Now, I'm sure I've just offended some devoted proponents of programmed instruction and CAI, so I hasten to add that I well recognize the fact that current instructional technologies have advanced to a point that their judicious application greatly enhances many learning experiences, but the constraints of computerized instruction inhere in computer programs' limited capacity to accurately interpret the stream of countless data points that human learners exude, and their limited flexibility in formulating reasonable courses of action that incorporate momentary situational variables. IBM's Big Blue's chess victory over Gary Kasparov notwithstanding. Of course, this is, quite remarkably, what human beings do with relative ease all the time.

When teaching is going well, assessment is ongoing. Skillful teachers are collecting assessment data throughout the course of instruction, because the incoming data stream is a necessary component of intelligent decision making concerning all aspects of teaching: when and how to convey information, assign tasks, and structure learning experiences that develop physical and intellectual skills effectively.

My inclusive view of assessment, which is at variance with many teachers' conceptualizations of assessments as culminating activities, encompasses all opportunities to learn about what students know, feel, and are able to do. Thinking of assessment as an ongoing part of teaching changes many things about how instruction is delivered. I have often heard teachers express that they "wonder how the students are going to do on this next exam." Hmmm. Why do they wonder? Why don't they know? The answer, of course, is that there have been few opportunities for these teachers to observe individual students doing the kinds of tasks that they will be required to do on the

examination, so there has been little opportunity to observe whether the students *can* do what will be asked of them when the assessment comes. This is a problem. This is a big problem. If the teacher has had few opportunities to observe students performing the kinds of activities that the assessment comprises, then the students probably have had little opportunity to determine the extent to which they, themselves, are able to perform the kinds of tasks that are on the examination.

A colleague of mine recently expressed a bit of consternation that her students had not performed as well on an exam as she and they had anticipated. Many students in the class had expressed to her that they "understood the material" and were "ready for the test," but they were unhappily mistaken. This raises an important question: What is the dependent measure for determining that one "understands the material" and is "ready for the test"? Now, every conscientious person who's good at anything knows that preparing for something important should include multiple opportunities to practice doing the upcoming important thing, whether that means delivering a paper at a conference, giving a speech, taking an audition, going through an interview, giving a recital. Most of us understand that the more opportunities we have to practice doing the important thing and the more precisely we can anticipate the circumstances surrounding the important thing, the better the important thing is going to go. But students often have much lower thresholds for "understanding" and "ready" than many practicing professionals, not because the students are inherently lazy or apathetic, but because of their inexperience.

Most novices judge the level of their own understanding based on their ability to understand *someone else's* presentation of ideas or to follow *someone else's* instructions to navigate a seemingly impenetrable problem; that is, "If I can understand the teacher's lucid, beautiful presentation of this knotty problem, then I understand it." (Many teachers are smiling right now because they know what's coming next.) OK, so now it's time for the student to explain this idea to someone else. It's test time. No longer following someone else's thinking, the student is now required to generate the linear progression of

thought independently, and here is where the breakdown occurs. Because the student has never generated the ideas on her own, because her primary experiences with the ideas have involved only listening to the teacher and the authors of the textbook explain the ideas, because she has thought about the ideas only in the contexts provided by others, she has had virtually no practice in doing what's being asked of her on the test.

I know that this explanation may seem astounding to some teachers, who may think to themselves, "I went over this material so carefully. I took them through all of the issues step by step. I gave them every opportunity to ask questions. I invited them to see me privately for extra help. How could they not understand after all of that?!" Because there were so few instances (perhaps none?) prior to the exam in which the students were required to generate explanations or practice [illegible] on their own and apply the information and skills in contexts other than the ones in which the material had originally been presented. That's why.

Many novice performers judge the level of their readiness to perform based on their ability to follow *someone else's* instructions to play a difficult passage; that is, "If I can follow the teacher's step-by-step instructions that lead me to a beautiful performance of this clumsy passage today in my lesson, then I can play this." OK, so now it's time for the student to perform the piece on his own in front of an audience. It's jury time. No longer following someone else's beautiful series of successive approximations, the student is now required to perform the piece independently. And again because the student has never performed the passage in this context on his own, because his primary experiences with playing the piece have been under the skillful guidance of his teacher in the studio or alone in a practice room, where stopping to fix mistakes was an inherent part of practicing and where graceful recovery from error was not a focus of attention, he has had virtually no practice in doing what's being asked on the jury.

This is further complicated by the fact that these same novices tend to evaluate their readiness based on their *best* performance trials; that is, their measure of how ready they are

to perform is based on their own best performance on their best day under ideal circumstances with little or no pressure. Of course, music performances seldom take place under ideal circumstances and with no pressure. One's readiness is best judged by the extent to which performances go well under circumstances that are as much like the real deal as possible—same performance space, with others listening, with the same the amount of warm-up time, even around the same time of day. Because most performances occur under less-than-ideal circumstances and with some measure of external or self-imposed pressure, having done it once, even really well, should give one little confidence that everything will come together when it's time to walk on stage. Of course, every expert performer understands this fact deeply. We know how many consistently correct repetitions are required before we consider ourselves "ready" for the performance. But many students grossly underestimate what it means to "be ready," and as a result of their under-preparation are often sorely disappointed by faulty performances that they had fully expected to go very well.

All learning requires active practice. Whether the subject matter involves physical skills or intellectual skills, practice is an essential component of developing understanding and skill. No practice, no understanding. As musicians we may think we grasp this better than most, which may be true to some extent, but even we seldom provide the kinds of active practice opportunities in our day-to-day instruction that afford students the practice necessary to internalize what we're attempting to teach them.

Returning to the story of my colleague who was disappointed with her students' test performance, how could she get to the exam and not have a pretty good idea of how everyone was going to do? How could the students be so clueless as to their own level of understanding and readiness for the test? Again, the answer centers on the infrequent opportunities in class to practice generating independent explanations and applying information and skills in novel contexts, all of which were required on the examination.

Now, some of my colleagues may argue that the practice I'm describing is the students' responsibility. After all, isn't that what studying and practicing are all about? That's why we build study spaces and practice rooms. Yes, of course this is what studying and practicing should be about. But if you haven't spent much time lately observing a student studying and practicing, I assure you that doing so will change your perceptions about your students and how they learn. Students need to *learn* to study and practice effectively and independently, but many have not yet learned to do so. I recognize that some teachers see effective studying and practice as volitional issues—if students wanted to study and practice well, they would. I heartily disagree. Students need to learn to study effectively, to practice effectively, to think effectively. So, when and where will they learn that? In class, with us. Not by our *telling* them what to do when they're alone in a practice room or in a carrel in the library, but by our leading them through the very activities that we expect them to do on their own in our absence. In other words, by practicing practicing in our presence.

What I'm suggesting here is that assessment be a part of every rehearsal and every class, that students have opportunities to demonstrate what they know and are able to do independently every time they meet with you. These opportunities need not be elaborate, time consuming, or burdensome, but they should be so frequent as to become a regular part of instruction. This means hearing students play alone in rehearsal (often), having students provide explanations in class (often), having students use the information and skills that they're working to master by applying them in ways that have not been explicitly taught (often).

Most arguments against what I've just recommended concern time constraints and the fear of intimidating students who are put on the spot by having to respond individually. Let's discuss the time issue first. First, it is not necessary that every student respond alone every day. You may select students to respond in class based on your perceptions of who can provide a good model for her classmates, who needs

opportunities to practice, and whom you need to learn more about. Once you are relieved of the burden of trying to get to every individual every day, this approach seems much more feasible.

Also, students may practice responding in small groups of two to four, which provides many more response opportunities in a given period of class time than is possible if individuals always respond in front of the entire class. Small groups are a multiplier of response opportunities. They provide not only more frequent opportunities for individual responses, they provide students opportunities to react to one another's ideas and evaluate one another's performances.

Regarding the concern that students who are called upon to respond alone in class will be intimidated by the experience: it is certainly true that students who have become accustomed to sitting quietly in class and responding only when they choose to volunteer may be more than a bit frightened by being called upon to explain an idea, solve a problem, or perform a passage alone in front of their classmates. But this fear is almost always a result of the infrequency of their opportunities to do so. If students are left to decide when and under what circumstances they will respond alone in class, many students will understandably opt to stay out of the glare of scrutiny by their classmates and the teacher. But because decisions about when and how often to respond play such an important role in learning, these decisions should not be left to students.

Permitting students to volunteer when they respond alone in class is a fundamentally bad idea, for reasons that are in many ways obvious. Think of who is likely to volunteer. For the most part, those who volunteer are either (1) confident students who are fairly certain that they know the answer or are confident that they can play the passage well, (2) students seeking the attention of the teacher or their classmates, and, most *in*frequently, (3) students who need help. What about students who understand so little that they fail to understand that they fail to understand? What about the students who need more practice opportunities but would rather not volunteer for reasons ranging from fear to apathy? What about the students who can play this passage beautifully but would rather not be singled out? These students and others may seldom if ever

volunteer to perform alone or answer a question in class, but there are many important reasons why they should.

A perceptive teacher who's collecting assessment data throughout the course of instruction understands these reasons and how they apply to each of her students. If she is in control of who speaks when, who answers which question, and who plays which passage or which portion of a passage, then the individual response opportunities can be arranged in deliberate ways that provide practice opportunities for those who need them, provide good student models of skilled performance, and provide structured illustrations of systematic error correction that leads to positive conclusions. These and other good reasons for students to respond in class support the notion that the timing of individual response opportunities can and should be under the control of the teacher.

In group instruction, the only way to gather assessment data that accurately reflect individual students' levels of understanding and skill is to observe *individual* students explaining, solving, playing, and singing. I'm often astonished by observing teachers who pose a question to a class, hear the question answered by a single student who volunteers, and proceed along with the lesson as if it has been adequately demonstrated that everyone understands what's going on. But only one student answered. How does the teacher know what the other students know? Of course, the teacher has no idea what the other students know, and the only way to find out is to collect data from the other students. This need not require that every student respond alone every day, nor does it require that students always speak or perform alone in class. Three students can explain an idea to one another in a way that illuminates the misunderstandings that may persist in a learner's thinking. Four students, each of whom is playing or singing a different part, can perform a passage together in a way that affords the students and the teacher a clear indication of who's sharp, who's late, who's out of balance, who's diction is muddy, or whose notes are too short.

Assessment drives instruction

Assessment drives instruction. Maybe it shouldn't. Perhaps it would be better if students, teachers, and the rest of us were not so overly concerned with how we're evaluated. Perhaps. But we are and it does.

Many teachers work to get students' attention off of assessment and on to the subject matter, in the hope of obviating the many questions that teachers hate to hear students ask, questions like, "Do we need to know this?" "Will this be on the test?" "How many lines do I have to play?" "How much of this are we responsible for?" Ugh. How did we get to this point? How did students develop such an obsession with their evaluations? Why did they come to believe that the only things deserving their attention and effort are things that will appear on a test or an audition?

The answer of course is that students *learned* to focus attention on evaluations, because in many ways, some of which are actually important, evaluation matters. By that I mean that the results of evaluation often have consequences that affect the future. In professional school admissions, GPAs matter. In scholarship deliberations, SAT and GRE scores matter. In music degree programs, jury grades matter. To many concerned parents, report cards matter. To students who want to sit first-chair, audition results matter.

None of this is inherently problematic. I'm certainly not suggesting that we need to teach students to adopt a cavalier attitude about their grades in school. Not at all. But the problem I wish to address here is that this heightened attention to and concern about important evaluations, for many (most) students, has generalized to *all* evaluations, regardless of their importance or the potential consequences associated with their results.

There are distinct disadvantages to students' being overly concerned with evaluation. First, it places the content of evaluations at the top of students' lists of priorities. What's important? Whatever it is that's going to be evaluated. Now, this would not be such a problem if the nature of most evaluations were such that what's tested were actually the most important, meaningful, and substantive aspects of the discipline studied. But most often this is not what's tested. Most

often, tests comprise many disconnected bits of information that "sample a student's knowledge and skills." The problem with the sampling idea is that most tests do not assess whether a learner can actually *apply* what he knows in useful, meaningful ways beyond the contexts in which the knowledge and skills are taught. Most tests require only that students remember and reproduce what they've been told or shown.

"I can define the terms volt, ohm, and ampere." OK, so explain how this electrical circuit works. "....I can define the terms volt, ohm, and ampere."

"I can name which of the 20 amino acids is encoded by each combination of nucleotide triplets." Fine. How does the ribosome translate mRNA into protein? "Did we go over that?"

"I can name the elements of one-point perspective." Terrific. How does the human perceptual apparatus interpret the information on a flat surface as having depth? "Huh?"

"I can name and aurally identify nine different forms that appear in common-practice music of the western canon: sonata form, rondo form, song form, minuet and trio...." Nice. So, what *is* form, exactly? "Hmmm....."

"I can identify the key name from any key signature." OK, so explain what tonality is. And, now that you know that this piece is in A major, how will that affect how you play the piece? Will you play it differently than you would have if you hadn't known the tonic? *Should* you play it differently now that you know that the tonic is A?

If you've never asked yourself these questions about key before and if you don't have an answer at the ready, then your thinking illustrates the necessity for more careful deliberation about the goals of instruction and the ways we go about evaluating their accomplishment. What *does* knowing the tonic have to do with how one plays a piece? If we can't answer this question beyond saying something like, "You know to play b-flat and e-flat," then we've got a lot more thinking to do, because of course, "You know to play b-flat and e-flat," just by looking at the key signature at the beginning of each line. You don't need to identify the tonic to know that. My point here is not that it is unimportant for student musicians to learn to identify tonic keys, but to illustrate the fact that teaching

practices and their attendant assessments are not always supported by clearly reasoned rationales.

The greatest problem with most current assessments is that they, the daily activities of instruction, and life beyond the classroom are all so utterly unalike. The obvious solution to this persistent and pervasive aspect of formal assessment is to design activities of daily instruction that more closely resemble life beyond the classroom and to design assessments that more closely resemble the activities of daily instruction. Meaningful assessments focus on what's most important about the discipline, and effective instruction includes frequent opportunities for students to actively practice applying what's learned.

Of course, all of this raises the question of what's important and meaningful, a question that is inextricably bound up with the notion of how learners will use the information and skills in the future, beyond school. Clarifying what's most important and deserving of instructional time begins with our thinking carefully about our expectations for the future. How do we expect learners to use and apply what they know in their lives beyond school? Understand that I'm talking not only about pragmatic applications of knowledge and skills that are work related. I'm referring to using information and skills in all of the ways that intelligent people use information and skills in their interactions with the world around them, at work, certainly, but also at home and at leisure, reading a book, writing a letter, interpreting the newspaper, making decisions about purchasing products, casting an informed vote in an election, making decisions about medical care, singing at a family gathering, reveling in figuring out how something works.

There is a tremendous difference between knowing stuff and applying what you know to accomplish meaningful goals. Being able to solve equations is a good thing, but if it's impossible to connect those equations to anything outside the world of mathematical abstraction, then being able to solve equations is merely academic, in the worst sense of the word, and pretty useless except insofar as it earns good grades on

math tests that require nothing more than solving equations. Knowing a wealth of facts about the history of Western music is nice, too, but if one is unable to weave together the events of history to construct a coherent view of the past, then knowledge of those facts can do little other than help score points on music history tests, beat your friends at Trivial Pursuit, and turn you into a bore at parties.

Class piano is perhaps one of the best worst examples of teaching knowledge and skills that seldom progress to their meaningful application. Class piano is a component of music instruction that has the twin disadvantages of being hated by most undergraduates and being mostly unproductive in improving their musicianship. I am seldom surprised by how poorly students who make Cs in class piano play the piano, but I continue to marvel at how poorly students who make As in class piano play the piano. How does this happen? How can students make such good grades and yet be so inept at the very skill the class is purported to teach? Well, because most assessments in class piano test skills that are of little value beyond the class piano exam.

Let's take scales, for example, and begin by asking the question that ought to be asked about anything that's taught, namely, What's the point? The point of scales is to develop fluency, flexibility, strength, and speed in the muscle movements that are required to play an instrument. Scales are exercises in coordination and calisthenics for the mind and for the body parts involved in tone production. And what is required to develop fluency, flexibility, strength, and speed through scales? Consistent, correct repetition of scales; that is, playing scales at speeds that are "fast enough" that the movements begin to become automatized to the point that the player no longer thinks of scales in terms of individual notes, but instead conceptualizes and performs scales and scale fragments as unitary constructs. No longer: note, note, note, cross thumb under, note, note…, but: scale. If a novice can play scales only very slowly, thinking deliberately about each fingering for each note, and that's as fast as she ever gets, then there is *no point* in performing scales. Think about this. A lot.

Of course, the rationale for teaching scales is that they form the basis of the pianist's technique and as such are a

requisite part of skill development. So far, so good. But how many of the students in class piano are going to continue to practice their scales after the requirement to do so is no longer present? Will they continue to practice scales until there is in fact a benefit from doing so? For most students in class piano (and you know who you are), the answer to this question is No. Thus, for those people who will not continue to practice scales to the point at which they begin to show a return on the investment of time they require, there is no point of practicing scales other than to pass the exit requirements for the course. (Having read the last sentence, a number of music teachers have hurled this book out the window.) Yes, but students *should* continue to practice scales until they reach a level of proficiency that is advantageous. Yes, but they *don't*. Yes, but they *should*. Yes, but they *don't*. Yes, but they *should*. Yes, *but they don't!*

So now what?

Will this be on the test?

It is undeniably true that assessments have tremendous potential to change behavior. Want to get students to pay more attention to what they're learning? Tell them you're going to give them a test over it. Want to get teachers' attention about what they're teaching? Tell them you're going to give their students a standardized test over it. Want to get administrators' attention about the quality of instruction in their school? Tell them you're going to give their students a standardized test and publish the results in the newspaper. Want to get superintendents' attention about the quality of instruction in their districts? Tell them you're going to give standardized tests and, based on the scores, you will either increase their districts' funding allocations or distribute portions of their extant funding to parents in the form of vouchers, which the parents can apply toward private school tuition.

You get the idea. Give tests. Define the test performance as a contingency for something that matters, and behavior often changes. There are many shopworn adages in education that express this same idea: What's tested is what's taught; assessment drives instruction; etc.

Many teachers and thinkers in education circles bemoan our increasing devotion to test performance as the coin of the realm in education. For many teachers, there is no more insulting pejorative than being accused of "teaching to the test," a phrase that connotes teaching disconnected bits of inert knowledge that contribute little to true understanding. Ouch. Why is this so? Why did "teaching to the test" become a slur to hurl at a disrespected colleague or callous administrator? Why is teaching to the test perceived to be such a bad thing? The answer is that most tests are pretty lousy at measuring a learner's understanding of the most important principles in a discipline. Most tests comprise items that require only that students remember facts or follow algorithmic paths to clear solutions to well-defined problems. But deep understanding requires more than that. Deep understanding involves the application of knowledge and skills in contexts that have not been taught explicitly.

The primary reason that most standardized tests do such a poor job of assessing the application of knowledge and skills in relation to what's most meaningful, interesting, and important about a discipline is that doing so is very expensive in terms of time, effort, and money. Psychometrics is a pretty sophisticated field, and skillful test designers have gotten very good at determining quite reliably and efficiently who knows what and who can do what. In the interest of efficiency and economy, however, most standardized assessments do not require that test takers apply information and skills in realistic (the hip education jargon for this is *authentic*) situations that require more than paper and pencil.

Assessing deeper understanding and the application of knowledge and skills requires a good deal of time and effort on the part of everyone involved: the test designers, the test proctors, the test takers, and the graders. The cost-benefit ratio is a real issue, not only for states considering how to implement a new system of public school accountability, but also for classroom teachers with limited amounts of instructional time.

The extent to which anyone is willing to devote their own time, effort, and energy to assessment is wholly determined by the real or perceived consequences of the assessment. In Texas, for example, regular instruction in grades 3, 4, 5, 6, 7, 8, 9, and 10 *stops* months before the administration of the Texas Assessment of Knowledge and Skills (TAKS), because a school's performance on the TAKS has personal, financial, and political ramifications. Performance on the TAKS *matters*. It matters to superintendents because their districts' publicized ratings of quality and budget allocations depend on it; it matters to principals because their jobs depend on it; it matters to teachers because there are often personal financial incentives to produce high scores; it matters to parents because, well, it's my kid, and, by the way, the value of my property is tied to the reputation of the schools in the district; it matters to students because one cannot receive a diploma without a passing score.

Those who rail against standardized testing as the nemesis of quality education do so not because they object to testing in principle, but because the tests themselves do not effectively measure the application of knowledge and skills in relation to what's most interesting and important about the disciplines tested. Understand that I am not an apologist for critics of standardized tests, and I'm sure that some whom I've lumped into my broad-brush description would disagree with my analysis of their motives. But I doubt that anyone who criticizes the excessive attention to testing and to students' test performance would be nearly as upset if the tests measured something that they felt was meaningful and important about what students know and are able to do. The fact is, unfortunately, that most standardized tests don't measure what's meaningful and important, their reliability and validity coefficients notwithstanding.

Consider an alternative where the tests do in fact measure what's most meaningful and important about the subject matter, in which tests require students to demonstrate the application of knowledge and skills in contexts other than those in which the knowledge and skills were initially taught. While providing a different kind of information to teachers about what students know and are able to do, these tests also convey a different kind of information *to students* in that the test content focuses on what's most important about the discipline.

Of course, just as assessments should embody the most important aspects of a discipline, so should the daily activities of instruction. I'm sure you see where I'm headed here. I'm suggesting that the distinction between the assessments and the substance of instruction day to day should be diminished to the point that the day-to-day activities of instruction closely resemble the assessments themselves. What goes on in class each day? Practicing the application of knowledge and skills that embody the most interesting and important principles in the discipline. What do the assessments comprise? The application of knowledge and skills that embody the most important aspects of the discipline. The assessments closely resemble the day-to-day instructional activities because they all embody the application of the core knowledge and skills in the discipline.

Once the goals for learning have been clearly identified in terms of how students will *demonstrate* what they know and are able to do, the instructional objectives and the assessment criteria have been composed in a single stroke. They are identical. And because these goals focus on the meaningful application of knowledge and skills beyond the contexts in which the skills are explicitly taught, then the daily activities of instruction can closely resemble the assessment activities. Thus, each day's instruction comprises opportunities to practice the meaningful application of knowledge and skills, and these practice opportunities are indistinguishable from the assessments themselves. Now there exists a seamless congruity among the goals, instruction, and assessment, all of which embody the most interesting, important, and meaningful aspects of the discipline. Practicing for the test is indistinguishable from learning the material. Think about this idea for a long while.

A vision of students as accomplished learners

The process of effective instruction begins with designing meaningful assessments, and at the core of meaningful assessment is a vision of students as accomplished learners. Planning, teaching, and evaluating the effectiveness of instruction are all predicated on a clear description of what

students will do to demonstrate that they've accomplished the goals set out for them. At the end of a given period of instruction—a day, a chapter, a unit, a six-weeks, a semester, a degree program—what do you expect your students to be like? What must they do to provide clear and convincing evidence of their competence?

The more vividly we can envision an exemplary student who has successfully accomplished the goals we set out, the more intelligently we can organize our instructional time and activities to facilitate the accomplishment of those goals by all of our students. This vivid image of students as accomplished learners at once defines the goals for instruction, the assessment criteria, and the substance of instructional activities, because the goals/assessments define how students will demonstrate their understanding and skills, and the instructional activities will be composed of structured opportunities to practice the goals/assessments.

There is no more effective means of providing structure, hierarchy, and priority to instruction than by composing the assessments before instruction begins. By this I mean actually sitting down before the start of the semester and writing the final exam or designing the culminating project assignment. The final exam defines the syllabus, because the final exam represents the embodiment of what you believe to be the most important goals of your class.

Teaching for the test in this way is entirely appropriate, because the test embodies what we care about most. The test is the tangible expression of the goals, which describe the *application* of knowledge and skills. Every superb teacher has a clear, vivid image of her students as accomplished, competent learners, engaging in the types of activities that closely resemble those that are likely to occur beyond school.

SONGLEADING PERFORMANCE CHECKLIST

Performance #: Date:

Guitarist: Score (.5 for each YES):

Song Title:

YES	NO	Posture: Sat up straight
YES	NO	Posture: Guitar in vertical plane
YES	NO	Began immediately when called (no nervous mannerisms)
YES	NO	Smiled
YES	NO	Gave correct starting pitch
YES	NO	Played appropriate introduction
YES	NO	Introduction: same tempo as song
YES	NO	Cued singers (e.g., "Ready, sing")
YES	NO	Looked at singers at start of song
YES	NO	Looked at singers at least twice during song
YES	NO	Looked at singers at end of song
YES	NO	Sang with singers during song (audibly and in tune)
YES	NO	Played correct chords
YES	NO	All strings vibrate with clear tone
YES	NO	Changed chords without pause
YES	NO N/A	If missed chord, continued without pause
YES	NO	Used appropriate right hand pattern
YES	NO	Played appropriate bass strings (strum or pluck)
YES	NO	Maintained steady pulse in right hand
YES	NO	Played loudly enough to be heard throughout

New Left Hand Technique or Chord:

New Right Hand Technique:

Specify positive aspects of this performance (at least four):

Specify one goal for improvement:

This evaluation accurately reflects the performance reviewed: _____
(Signature)

The fundamental principles that form the basis of intelligent assessment can perhaps best be conveyed through an example. Pictured above is a checklist I use for evaluating guitar playing and song leading. I adapted this checklist some 20 years ago from a dissertation project conducted by a

graduate school classmate of mine.[1] Look it over for a minute or so and consider carefully the items on the list. Note that the checklist comprises 20 verb phrases, which, taken together, outline what a confident, competent, guitar-playing song leader does when she's performing well. The composite of all the individual items is essentially a concise description of a confident, competent, guitar-playing song leader in action. If guitar-playing song leaders do everything on the list, they look and sound pretty good.

How was the list generated? By beginning with *a vision of students as accomplished learners*—which in this case means a vivid image of a competent, confident, guitar-playing song leader—and constructing a clear and concise description of what that competent, confident, guitar-playing song leader looks and sounds like. The resulting list comprises all of the essential components of good guitar playing and song leading.

Note that the checklist does not define the difficulty level of the piece performed. In this sense it is context-independent, because the skills described on the list are applicable whether the performer being evaluated is playing his first two-chord song or is the opening act for Asleep at the Wheel. All competent, confident, guitar-playing song leaders do (or should do) all of the things on the list every time they play, throughout their musical journeys from novice to expert. Using a single list to describe the fundamental skills of musicianship, or any other skill for that matter, over an entire course of study recognizes that skill acquisition is a process of developing habit strength through consistent correct repetition over time.

Nor does the assessment include scaled ratings for each of the variables on the list. Instead, each item is assessed with a dichotomous Yes or No. I will explain this briefly here only by saying that applying an evaluation scale to each item on the list unnecessarily complicates the assessment and provides no useful information for the students. Making judgments about what rating to assign requires time and thought but contributes little of value. "Well, he sort of smiled and he played most of the chords, and he only paused once." So, out of a possible 5

[1] Furman, C. E. (1985). *Behavior checklists and videotapes versus standard instructor feedback in the development of a music teaching competency*. Unpublished doctoral dissertation. The Florida State University.

points for each item, is that a 2 and a 4 and a 3? If it is, what does the average mean? Is it to be interpreted as an overall score? Is this a reasonable estimation of the performer's skills? You see the growing complication here, a complication that brings the added disadvantage of being pretty useless, except as a basis for arguing about grades.

But shouldn't the student who played most of the correct chords get some credit for that? Does missing one chord really warrant a No? Well, the student didn't play the correct chords, so a No for Played Correct Chords is appropriate in this case. The reason it's appropriate to assign a No to someone who missed only one chord is that Played Correct Chords is only 1 of 20 criteria, and it contributes only 1/20th of the student's score. The effect of missing any one criterion is of little consequence in terms of the student's grade, because the relative weight of each aspect of performance is built into the assessment criteria.

Note some of the important principles at work in this approach to assessment. The first is that students know at the outset (1) what is important about what they're learning, (2) what is expected of them in terms of their own performance, and (3) how their meeting these expectations will contribute to their grade in the course.

The second point is that the assessment criteria remain the same throughout the course of study. Why? Because the assessment criteria are not based on course content (in this case, repertoire), but are based on the application of intellectual and physical skills. The fundamental skills don't change as one becomes more and more expert in a discipline, but the contexts (repertoire) in which the skills are applied become more and more sophisticated and demanding.

The third important point about this form of assessment is that the assessment criteria focus on the application of knowledge and skills in context—in particular, in a context that is very much like the extracurricular contexts in which the knowledge and skills are likely to be applied long after the course is over.

The fourth point is that the items are constructed so as to apportion weight appropriately, and as a result the students are less likely to ascribe undue importance to aspects of their

performance that are not as consequential as they might be perceived to be by a novice.

The fifth point is that the context for the assessment requires students to demonstrate all of the skills on the list from the very beginning of their experience as learners. Of course, it's no big deal to smile (at least for most), begin immediately when called upon, sing loudly enough to be heard, or use the appropriate bass strings. The hard part is doing all of these things at the same time while your performing in front of your peers and your teacher. And it's the simultaneous application of skills that keeps people from becoming confident, competent, guitar-playing song leaders, excellent point guards, successful personnel managers, insightful researchers, and skillful teachers.

The final point is that this set of criteria accurately expresses what I think is most important for the students to learn from this experience. What do I care about as your teacher? Look at the variables that contribute to your grade in this class. They express what I care most about your learning from this experience. Speaking to colleagues in higher education, I ask from time to time whether they would feel comfortable if their own values regarding their disciplines were inferred only from the formal assessments that their students performed. In other words, if a stranger knew you only on the basis of a review of your methods of student assessment, would that person have an accurate picture of what you care most about or what you think is most interesting in your discipline? Many of my colleagues haven't thought much about this question, but upon reflection, they have expressed to me that the means by which they assess their students often is only remotely related to what they care about most or what they find most interesting.

The rationales offered for this disconnect between what teachers value and what they assess often involve extended discussions about having to "cover" certain information or repertoire so that students are "exposed" to the basic components of the discipline. These rationales do not hold up under the scrutiny that questions breadth versus depth. Although you may be able to "cover" the information or skills listed in the course description, can students apply the

information and skills in ways that are useful beyond the course? Is the basis of your assessments of student performance their ability to remember and reproduce what they've been told or shown, or does it involve their using what they know and are able to do to solve novel problems and perform skills that were not explicitly taught?

Tests teach

Instructional goals are meaningful only insofar as their accomplishment can be demonstrated by a learner. Statements of intention, that "students will learn to appreciate music," or that "students will learn to think creatively," or that "students will learn to use higher order thinking skills," for example, are meaningless without adequate descriptions of what students will do to demonstrate that they "appreciate music," "think creatively," and "use higher order thinking skills." Absent explicit definitions of the standards of evidence required for documenting appreciation, creativity, and higher order thinking, goal statements like the ones above contribute little to quality instruction.

Goal statements that do include explicit descriptions of what students will do to demonstrate that they've accomplished the goals are, in fact, descriptions of assessment criteria. Statements that "students will choose to listen to three different instrumental selections from the repertoire of the Western art music tradition," or that "students will improvise a three-note melody over a 12-bar blues progression, using the tonic, supertonic, and mediant pitches in a comfortable key," or that "students will explain the relationship between Brahms's symphonies and the symphonies of Beethoven," describe both the goals of instruction and the tasks that students will perform to demonstrate their accomplishment of the goals.

Given the fact that in this way of thinking the goals of instruction and the criteria for assessment are essentially one and the same, how should you begin to define what these goals/assessment criteria should be? Our discipline, like all disciplines, comprises innumerable bits of information and countless component skills. What should serve as the guiding principle in sorting through all of the stuff that can be taught

under the heading of Music and defining a set of tangible goals and assessment criteria? What shall we require of students to demonstrate that they have learned what we set out to teach? What will we require as evidence that students have learned?

The answer is first to imagine a competent person who demonstrates all of the skills and knowledge that we're intending to teach and then to define with some precision what that person is like. What does she do? What does she know? What attitudes does she convey? This image supplies us with a vision of our students as accomplished learners, as individuals who demonstrate competence in the fundamental aspects of the discipline. A precise description of this vision defines the assessment criteria. How will we evaluate students' knowledge and skills? By having them do the very things that competent professionals do, but with contextual limitations that are appropriately gauged to learners' levels of experience and expertise.

The difficulties surrounding the design and implementation of assessment can be organized around three dichotomies: skills-content, breadth-depth, and frequency-magnitude.

Skills versus content (Doing stuff versus knowing stuff)

Genuine competence in any discipline is fundamentally based on skills: reading skills, listening skills, reasoning skills, communication skills, social skills. Competence is much more than knowing stuff (content); competence entails doing stuff with what you know, using what you know to reason, solve problems, pose questions, play music, read, discuss, write. Thus it is advantageous for us as teachers to think about the assessment of learning not in terms of what students know, but instead in terms of how students use what they know to accomplish goals. We need not worry that by emphasizing skills we'll shortchange the knowing-stuff, because assessment of the knowing-stuff is subsumed within the assessment of doing-stuff. Knowing that written E on the B-flat clarinet sounds concert D, for example, is inert knowledge. Writing a woodwind quintet arrangement of a four measure melody puts this knowledge into action, applying the knowledge of

transposition to accomplish the goal of writing for woodwinds. You have to know stuff (e.g., that B-flat clarinet parts sound a whole step lower than written) to do the doing stuff, so by assessing the doing stuff, the knowing stuff is also assessed.

This focus on skills rather than content is a liberating way to think about goals and assessments, because it alleviates concern over covering material. Teachers who feel pressured by time constraints often move quickly through skills and ideas in the interest of getting through what "must be covered," even though such a rapid pace affords students few opportunities to practice and refine skills. But the development of skills requires frequent opportunities to practice applying the skills in a variety of contexts. No practice, no skill development. Once the decision is made to forgo allegiance to covering the material (getting through the content), then more time may be spent demonstrating, analyzing, practicing, and refining skills (applying the content to accomplish goals). Focusing on skill development rather than content coverage not only reorients teachers' priorities about the use of precious instructional time but also reorients students' perceptions of what's important, what's deserving of their attention and effort, and what it means to be competent.

Most of the intellectual and physical skills we hope our students will acquire through the course of instruction are not context dependent. Our goal is not that students play this here whole step in tune, for example. Our goal is that students play whole steps in tune, irrespective of where they occur in the pieces they play and sing. Our goal is not that students maintain a steady tempo in this here etude. Our goal is that students learn to maintain steady tempos generally. In other words, our goal is not that students will play or sing a given repertoire (content) but that our students will develop the habits of good musicianship (skills). We may approach this goal through any number of exercises, etudes, and pieces, but irrespective of what's being played or sung at the moment, the goals of musicianship—the skills—are unchanging throughout the course of instruction from the first days of producing a tone.

From this point of view the repertoire performed in a music assessment, for example, is not nearly as important as the

quality of the performance. Think about what can happen to students' thinking and priorities when there is no credit for performing very difficult repertoire rather poorly. It's not *what* you play or sing, it's *how well* you play or sing. Sloppy Paganini doesn't win over beautiful Handel, because the criteria now focus on the quality of performance: the beauty of tone, accuracy of intonation, rhythmic precision, expressiveness. The assessment emphasizes that it's the demonstration of refined skills that matter, not the difficulty of the context in which the skills are applied.

Depth versus breadth

The same teachers who make the unfortunate decision to base their instructional goals on content coverage rather than on the development of physical or intellectual skills seldom devote sufficient time and attention to any one idea, principle, or skill, so that students come to understand it deeply or perform it fluently. In the competition between breath (coverage) and depth (competence), breadth often wins. I've heard many teachers insist that they "must get through this material by the end of the semester," as if there is some external force directing them to make this regrettable choice. Although I certainly acknowledge the existence of curriculum guides, textbooks, standardized tests, principals, chairmen, area supervisors, deans, and parents, the assertion that one or more of these forces requires that teachers blow through material faster than students can take it in is demonstrably untrue.

Instructional time is a zero sum game, and increasing the time devoted to any one topic or skill necessarily diminishes the time available for other topics. But it is generally impossible to cover a great deal of material and develop depth of understanding and fluent skills. There is simply insufficient time to do both. When more time is devoted to a few fundamental skills and a few underlying principles of the discipline and students have many opportunities to practice applying these skills and principles in a variety of contexts, increased competence, fluency, and confidence are the results. The breadth versus depth question is no contest at all. Depth trumps breadth every time. If we expect students to come to value excellence, then we must afford them opportunities to cultivate, nurture, and practice excellence. This represents a

considerable investment of time and effort. It doesn't come quickly and it doesn't come cheap, but its value is unassailable. We cultivate excellence by expecting students to demonstrate excellence—depth of understanding and fluency of skills—and by providing numerous opportunities for them to do so.

If we expect students to learn to value deep understanding and intellectual and physical fluency, then we should be certain that all assessments, formal and informal, convey that value. If assessments require only that students repeat what they've been told or imitate what they've been shown, it is unlikely that they will learn to value and invest their time, effort, and energy in their own intellectual or physical skills. This is especially true if assessments are designed in ways that set up students to do poorly. If students are performing repertoire at the brink of their current technical capacity, for example, what is the likelihood that they will play beautifully, expressively, artistically? What if instead of deciding a priori that students must play or sing repertoire at a given level of difficulty, we insist that students play or sing only repertoire that can be performed beautifully, expressively, fluently? In the first instance, everyone performs the same repertoire. Some sound wonderful; others stink. In the second instance, everyone performs repertoire that they can perform beautifully. Some play or sing some very difficult pieces; others perform very simple tunes. But *everyone plays and sings beautifully*. And through repeated experiences like this, the habits of competent performance develop, because all students play and sing beautifully repeatedly.

I realize that some may interpret the preceding as my suggesting that we lower our standards in order for all students to succeed. Quite the contrary. I am advocating for very high standards, because, in the second instance I described above, everyone must sing and play beautifully, a goal that is accomplished by limiting what students are permitted to perform to those pieces that they can perform beautifully. Note however that my high standards are based not on difficulty of repertoire but on quality of performance. Although this view of standards seems to me axiomatic, I have heard it articulated by few devotees of the so-called standards movement. Most often when I hear people argue for higher standards, they're arguing

for harder tasks. Of course, this is the cheap and easy way to raise standards: make the assessments harder. It's cheap and easy because it doesn't require that you do anything other than make the test more difficult or raise the passing score. It doesn't require that you teach better, spend more time, or invest more money in materials and equipment.

The success of this type of standards raising is predicated on the naïve belief that improving student performance requires only that the consequences of success and failure increase in magnitude, a belief that curiously persists among many of my Republican friends in spite of the glaring failure of states to improve education by simply raising the bar, of which the New York State Regents Exam is only one of many stunning examples. I certainly do not deny that our expectations can greatly influence what students accomplish, but telling a student who's never learned to think linearly that he'll be held back if he can't perform quadratic equations or write a persuasive essay will do little other than create frustration, anxiety, discouragement, resentment, and no small measure of anger. It certainly won't teach him to think linearly, solve equations skillfully, and write lucid and cogent essays. It will take many, well structured, repeated experiences to do that.

Frequency versus magnitude

Perhaps because so many teachers and students have come to view assessment as an onerous big deal, assessments tend to happen infrequently. In some college lecture classes, for example, students' grades are determined only by a mid-term and a final exam, with few experiences taking place during the other 44 class days that resemble the activities the examinations comprise. What's immediately apparent in this extreme way of doing things is that these "grade producing situations," as they're sometimes called in the education jargon, take on considerable weight in the minds of students. These tests become a big deal both because so much is riding on them and because of their infrequency.

To the extent that such grade producing situations as midterms, finals, juries, and auditions motivate students to get with it and study or practice in earnest, they focus students' attention on the content of the assessments, and if the assessments in fact tap the most important and substantive

aspects of the discipline, then this heightened attention to what's most important is not a bad thing. Unfortunately, however, many of these assessments fail to tap the most important and substantive aspects of the discipline, while at the same time creating stress and frustration in students who attempt to cram too much stuff into too little time. Of course, this way of learning is antithetical to systematic habit building. In the absence of sufficient opportunities for consistent, correct repetition over time, even the knowledge that may be remembered, the principles understood, or the skills executed for the test will remain "fragile" in the sense that they will not likely persist much beyond the assessment itself. To understand deeply and to think and move fluently requires active, consistent practice.

I recognize that many of my colleagues argue that the consistent practice and steady effort applied toward understanding and skill are the students' responsibility. Our job as teachers is to convey information, to explain, to demonstrate. It's the students' job to study, practice, and make what we tell them a part of what they think and how they behave. This is a very limited view of the teacher's role in the learning process. To teach effectively we must do much more than convey information. Teaching is much more than telling. The heart of teaching isn't the telling part but the systematic structuring of learning experiences (thinking experiences and doing experiences) that guide students through the acquisition and development of intellectual and physical skills. You can't talk someone into competence. You have to *do* someone into competence.

Assessments contribute to this process when they are generally high in frequency and low in magnitude. Frequent assessments provide opportunities for students to practice applying what they know and demonstrating what they can do, providing important information to teachers and students alike, but also regularizing the act of performing the intellectual and physical skills that everyone is working to master. The increased frequency and familiarity with the tasks associated with the assessment—which ideally are the very tasks that embody the most important aspects of the discipline—render

assessment just another part of the learning process. No panic. No sweat. This happens all the time.

The increased frequency also reduces the perceived consequences of error, and this is perhaps the greatest contribution that frequent assessment opportunities make to effective instruction. Students' perceptions of the consequences of error are generally out of whack. All of us have observed or experienced bright, capable, mostly well-adjusted human beings go to pieces before our eyes while arguing over a half-point for a partially correct answer—a half-point that doesn't affect their final grade. I realize this is so common as to be considered an inevitable part of schooling, but it's weird. It's very weird. And it illustrates students' and teachers' lack of perspective about assessment and grading. Think about this carefully.

All intelligent, skillful professionals recognize that error is an inevitable, necessary, and even productive part of thinking and learning. It is virtually impossible to develop genuine competence or to accomplish anything important, for that matter, without sometimes making mistakes, pursuing unproductive paths, or misinterpreting data. Making mistakes is an accepted part of learning new things just about everywhere except in school. In school mistakes cost you, because mistakes lower your grade. And the more *infrequent* the assessment opportunities, the more costly the mistakes.

Understand that we're trying to teach students to adopt a very subtle point of view: don't make mistakes if you can avoid them, but accept the fact that mistakes will happen because they are an inevitable part of learning; try not to make mistakes, but it's OK if you do. The difficulty of acquiring this nuanced way of thinking about learning is evidenced by its rarity among our students. To most of them, mistakes are bad, period. This position is reinforced to the extent that being wrong lowers your grade in a consequential way. When we provide few graded assessment opportunities, we build in *disincentives* for students to comfortably recognize their own errors and accept criticism regarding their work. They can't afford to. Doing so would lower their grade. If error is always associated with a consequential diminution of a learner's grade, what intelligent, motivated student would come to accept error

as a natural part of the learning process and appreciatively accept our criticism as a welcome bit of new information that will contribute to their intellectual growth and skill development?

Assessment can and should be a regular part of the process of instruction, even to the extent that it becomes indistinguishable from the process of instruction. If we accept the premise that learning requires the active application of knowledge and skills and the complementary premise that assessments should embody the meaningful application of knowledge and skills, then there should be no impediment to making assessment an ongoing part of teaching and learning. We can accomplish this by creating more frequent assessments that are lower in magnitude in the sense that they are less time consuming and in the sense that each contributes little to the final grade. What often deters teachers from scheduling frequent assessments is the erroneous belief that every student has to be evaluated at the same time. They don't. Or that everyone has to do exactly the same thing on the evaluation. They don't. When assessment is woven into the fabric of instruction, the timing of assessments may vary according to the circumstances.

If all of this seems odd to you, it is only because we have become so accustomed to doing things as they've been done for many, many years. One of the greatest impediments to our thinking creatively about assessment, or about any other aspect of teaching and learning, is that we've all gone to school ourselves. And through our school experiences, we've developed habits of thought that do not necessarily lead to our focusing on the core principles of our discipline. This is not unique to music instruction. It pervades public education in every field of study. We often learn disconnected minutiae about math or science or social studies, wondering how anyone could possibly find this stuff interesting.

Teaching well begins with our formulating a vision of students as accomplished, literate, inquisitive, skilled, thinking musicians. If we take the time to describe an accomplished learner with some precision and if that description is

sufficiently explicit, then we have defined simultaneously not only the assessment criteria but the goals of instruction and the nature of the daily class activities as well. All in a single stroke.

This way of thinking about assessment shifts the emphasis away from the *activities* that take place in music class and the *repertoire* that is performed in choir, band, or orchestra, and toward the fundamental skills of intelligent, literate, musicianship that all of us intend for our students to learn. The point of our instructional practice is not, after all, to teach students to perform a particular piece or to participate in a given activity. The music repertoire and instructional activities are only the experiences through which we develop knowledge and skills that will be applicable, meaningful, and useful beyond school. It is these life-long musical and intellectual skills that teachers are working hard to develop, because they form the basis of what students will take with them when they leave school.