

An NCME Instructional Module on

High Quality Classroom Assessment: What Does It Really Mean?

Richard J. Stiggins

Northwest Regional Educational Laboratory

Teachers who gather accurate information about student achievement through the use of sound classroom assessment contribute to effective teaching and learning. On the other hand, those who fail to understand and apply the rules of evidence for sound assessment risk doing great harm to students. Thus, all teachers must understand the differences between sound and unsound assessments. This module is designed to promote that understanding. It examines the many users and uses of classroom assessment, the wide range of achievement targets to be assessed, the array of assessment methods teachers use, and the importance of marrying targets and methods in ways that promote

sound assessment. Four key attributes of sound assessment are presented for the teachers to apply in their own classroom assessment environments.

The quality of instruction is a function of teachers' understanding of the strengths and weaknesses of their students. The depth of that understanding, in turn, hinges on the quality of teachers' assessments of student achievement. Thus, sound instruction requires the sound classroom-level assessment of student achievement.

Research suggests that teachers spend as much as one-third to one-half of their available professional time involved in assessment-related activities. They are continually making decisions about how to interact with their students, and those decisions are based in part on information they have gathered about their students through classroom assessment (Stiggins & Conklin, 1992). And yet, other research reveals the troubling paradox that teachers receive virtually no relevant training in classroom assessment as part of their professional preparation, undergraduate or graduate, preservice or in-service (Schafer & Lissitz, 1987). Teachers typically are asked to conduct all of the assessments that drive the day-to-day instructional process—the assessments that inform the decisions that control student learning. Yet, usually teachers have not been given the tools nor had a chance to develop the expertise for this responsibility. This situation explains why so many teachers are uncertain and uneasy about their assessments of student learning.

This overview has been written to provide classroom teachers and those who offer direct support for classroom instruction with the basic assessment guidance and tools needed to track student achievement in the classroom on a day-to-day, week-to-week, and term-to-term basis. However, it is written only to represent a starting point in a teacher's professional development in assessment. Upon completing this unit of instruction, a teacher will have a clearer sense of his or her additional assessment training needs.

In this introduction, we will analyze the elements of the classroom assessment process. We will explore why teachers assess, what they assess, how they assess, and what can go wrong when teachers try to assess. That is, first we will examine the full range of users and uses of classroom assessment. Second, we will consider the full range of achievement targets that teachers must assess. Third, we will list and describe the broad range of assessment methods

Editorial Note: The inclusion of this module in the ITEMS series signals a significant broadening of the ITEMS focus. The vast majority of modules produced to date have dealt with relatively technical issues that are important in measurement instruction but that typically are not covered in depth in graduate-level measurements texts. Thus, ITEMS have served as valuable supplements to texts in the measurement training context. In the future, modules will continue to serve this purpose. However, beginning with this module, ITEMS will also provide units of assessment instruction designed especially for inclusion in undergraduate and graduate teacher and school administrator training contexts. These modules will be written to convey important assessment, evaluation, and grading concepts and strategies in nontechnical language, as modeled in this unit. The purpose of these modules will be to help practitioners integrate assessment into the day-to-day teaching and learning process. NCME members involved in teacher and/or administrator training, whether preservice or in-service, are encouraged to reproduce and integrate these units into their repertoire of training tools. Members interested in contributing practitioner oriented modules to the ITEMS series are urged to contact the editor as soon as possible at the following address: Richard J. Stiggins, ITEMS Editor, Northwest Regional Educational Laboratory, 101 S.W. Main, Suite 500, Portland, Oregon 97204.

Series Information

ITEMS is a series of units designed to facilitate instruction in educational measurement. These units are published by the National Council on Measurement in Education. This module may be photocopied without permission if reproduced in its entirety and used for instructional purposes.

available to teachers for tracking student growth. And finally, we will consider the various roadblocks to sound assessment and derive from that consideration a meaning of quality assessment that any teacher can apply to any assessment in any context to see if students are being well served from an assessment point of view.

The purpose of this analysis is to cause the teacher to reflect on his or her own classroom assessment practices and to evaluate both the soundness of those practices and the impact of assessment on students in the classroom. To encourage such professional reflection, throughout the guide, brief inserts offer "issues to ponder." We hope you will take time to do so.

Users and Uses of Classroom Assessment

All who have a vested interest in student learning use the results of classroom assessments in important ways. Those with the most direct interest include teachers, students, and parents. However, building and district administrators and school board members may also use these results. No single assessment or assessment format can serve all users and uses. It will become clear as this discussion unfolds that one important key to sound classroom assessment is a clear sense of the purpose for any particular assessment.

Teachers as users. Teachers represent the focal point of classroom assessment environments. They devise, administer, and use assessments to serve at least three purposes: to inform specific decisions, to instruct, and to control student behavior. Each of these categories has a number of more specific uses contained within it. For example, teachers make many of the decisions that make instruction work when they diagnose student needs (individually and in groups), group students for instruction, grade student performance, and/or select students for access to special services. Each of these decisions bears a direct relationship to the quality of instruction. If any of these decisions is based on misinformation about student achievements, instruction is likely to be ineffective.

But teachers are more than decision makers; they also instruct. And so it is with assessments. Assessments don't just inform decisions; they are also used to teach. Teachers often use assessments to inform students about their expectations—to let students know what kind of performance is required to be successful. Teachers also use practice tests to help students learn to hit those targets, that is, to help students internalize the required knowledge and skills. And teachers use assessments to provide students with information about performance so students can make some of their own decisions about school and learning (more about that later). Clearly then, assessments are also teaching tools.

Teachers also use assessments as behavior-control and classroom-management tools. They use the anticipation of evaluation and judgment to encourage studying. Further, for better or worse, they use the threat of tests and grades as a source of power to keep students in line. When students behave in a socially accepted way, teacher evaluations of achievement (a student attribute that is ostensibly independent of social behavior) can misrepresent actual achievement, such as through higher grades. And when students misbehave—whether they learned more or less—teacher evaluations can also misrepresent true achievement, such as through lower grades. Thus, teachers sometimes use the sheer power of assessment to force specific behaviors, attitudes, values, etc. This represents a little acknowledged, but very important, reality of classroom assessment that carries with it both legal and ethical requirements of teachers.

Students as users of assessment. Student uses of classroom assessment results are as important for student well-being as are teacher uses. Like teachers, students are decision makers who rely on teachers' classroom assessment results to inform

their decisions. Just think about the role that the following decisions play in determining the success of schooling: a) students decide on what personal academic expectations to set for themselves based on teachers' evaluation of prior achievement; b) students decide who controls their academic well-being based on the relationship they see between how hard they work and how much the teachers' classroom assessments indicate they have learned; and c) students decide if, what, when, where, and with whom to study based on feedback from classroom assessments. As with teachers' decisions, if any of these student decisions is based on misinformation about achievement, poor decisions will be made by students and learning is likely to suffer.

Also, like teachers, students are more than decision makers. Students study—they practice using assessments (such as assignments) to hone their knowledge and skills. They use practice assessments to evaluate their own achievement and to adjust, so as to internalize more of the knowledge and skills required by the teacher.

Parents as users of classroom assessment. Over and above the teacher and students, others use the results of classroom assessments. For instance, parents set expectations, plan educational resources, set home study environments, and assist with instruction based on their understanding of the ongoing achievement of their children. That understanding comes from feedback from teachers on day-to-day achievement communicated on homework, tests, report cards, and other messages sent home. The decisions parents make about how they will relate to the entire schooling experience of their children obviously are critical to their children's academic well-being. Parents make those decisions based on (a) the match between their expectations for their children and their children's apparent achievement, and (b) their perceptions of the "teachability" of their children. These perceptions arise from ongoing parental interpretation of teachers' classroom assessments.

Issues to ponder. Given all of these critical roles that assessments play in the classroom, what do you think would happen if a teacher planned an assessment without reflecting on how it was to be used? What might happen if a teacher was insensitive to how students are likely to use a particular piece of assessment feedback? What might happen if various assessment purposes became confused, such as when an assessment meant for diagnosis or practice became woven into a student's report card grade?

Assessing Achievement Targets

In addition to a clear understanding of the purpose for assessment, another key to sound assessment is a clear vision on the part of the teacher of the achievement targets he or she wants students to hit. It is impossible, for example, for a teacher to assess a student's level of writing proficiency if that teacher does not clearly understand the attributes of good writing.

In most educational contexts, teachers hold expectations that students will hit one or more of the following kinds of achievement targets:

1. Knowledge—Teachers expect students to master substantive, verbalizable subject matter knowledge (e.g., science or history facts).

2. Thinking—They often expect specific higher order cognitive operations or problem-solving skills to be demonstrated by students using the knowledge they have mastered (e.g., drawing inferences).

3. Behaviors—They expect achievement-related skills to be exhibited by the student in specific academic contexts (e.g., giving a speech).

4. Products—They expect achievement-related products to be created by the student, and they want these products to possess certain specific attributes (e.g., a research report).

Table 1
Classroom Assessment Options

Method	Objective form	Subjective form
Paper and pencil	<ul style="list-style-type: none"> ● Multiple-choice ● True-false ● Fill-in ● Matching ● Forced-choice questionnaire (affect^a) 	<ul style="list-style-type: none"> ● Essay ● Open-ended questionnaire (affect^a)
Performance assessment	<ul style="list-style-type: none"> ● Checklist of attributes present or absent in performance 	<ul style="list-style-type: none"> ● Rating scales ● Anecdotal descriptions of performance
Personal communication	<ul style="list-style-type: none"> ● Instructionally relevant questions answered right or wrong ● Oral exams—correct or incorrect 	<ul style="list-style-type: none"> ● Interviews, conferences, or discussions where depth of achievement or affect is explored ● Oral exams—open-ended

^aUsed to assess various affective characteristics such as attitudes, values, interests, etc.

5. Affect—They hope students will develop certain dispositions, academic self-concepts, attitudes, and values.

For teachers to assess and teach effectively, they must possess a clear and highly refined vision of each of the targets they hold as valuable for their students. They must know in explicit terms what students are supposed to know, think, and be able to create when instruction is completed. And further, they must know how those targets translate into assessments that make sense (more about this in the next section).

Issues to ponder. How does a teacher assess thinking without a clear, verbalizable definition of what it means to think and solve problems? Who is more likely to succeed at hitting an achievement target: a student given a clear vision of that target by the teacher in advance of instruction or a student left to figure out the target on his or her own? Do circumstances arise where teachers might be aiming for two or more of the targets listed above simultaneously? If so, what does this suggest about the need for teachers, students, and parents to understand and be able to communicate about achievement targets? What proportion of achievement targets do you value as outcomes for students that fit into each of the categories listed above?

Assessment Tools

In addition to a clear sense of purpose and a highly refined vision of the target, another key to assessment in the classroom is a thorough sense of the range of assessment options available for classroom assessment. While we often think of assessments as collections of multiple-choice items, classroom assessments include far more. Certainly paper-and-pencil instruments continue to have a valuable role to play in classrooms, but so do performance assessments (teachers' observations and judgments) and personal communications with students. Further, any of these three types of assessment can be cast as either an objective or a subjective assessment. This distinction is crucial because it allows us to describe the extremely broad array of assessment alternatives teachers have at their disposal to assess classroom achievement targets. Objective assessments judge the quality of the response as acceptable or not, right or wrong, all or none—no room for judgment. Subjective assessments, on the other hand, rely on the professional judgment of the teacher to evaluate the student's response along a continuum of performance.

Table 1 provides examples of the three basic types of assessment in objective and subjective forms. The list of possibilities is long indeed. And then, on top of this long list of options, we must realize that both paper-and-pencil tests and performance assessment can come from any of three different origins: teacher developed, text embedded, and standardized. Further, all of these can be given to students in at least two forms: as test/quizzes or as assignments. And still further, these assessments can be used by teachers to assess students or by students to assess themselves. These factors assembled in all possible combinations provide the teacher with an almost infinite array of different forms of classroom assessment.

The Meaning of Quality: Using Tools in Context

So how do we decide which forms of assessment to use in any particular context? First and foremost, it's a matter of the achievement target to be assessed. Certain targets match up best with certain assessment methods. For instance, assessments of mastery of knowledge might be done most effectively and efficiently with paper-and-pencil tests, while behaviors might link best to performance assessment. Consider the possible matches between method and target depicted in Table 2. These only represent some of the possible combinations. Clearly, without a vision of the valued outcomes, we cannot make an informed selection of a proper assessment method.

Fitting the context. Other contextual matters to consider in selecting an assessment method include purpose, students, and time considerations. Depending on the context, the user is more or less able to take advantage of the strengths of a particular method and/or overcome its limitations. Examples of some of these strengths and limitations are listed in Table 3.

Using assessment methods well. To summarize, one key to successful assessment is a clear vision of the achievement target. Another is a clear sense of the purpose for assessing. And a third is a complete understanding of the full range of assessment options. But once a particular method is selected, the user must avoid certain pitfalls in order to achieve sound application of the selected method. Those who assess well know these pitfalls and how to bridge them.

Pitfalls in assessment can cause us to mismeasure student achievement. They can interfere with our ability to know what students have learned. They can only be avoided if we recognize

and confront them. For example, when we're testing via objective paper-and-pencil methods, we need to sample the knowledge and thinking skills with enough items to be sure we can generalize to all the other questions we wanted to ask but didn't have time for. Thus we need a representative sample that is big enough, but not too big. There are no rigid rules for determining how many is enough. Assessors know they have asked a student enough questions when they can accurately predict how a student would do on the next one, if it were asked. It's a sense, a feeling. We don't want to waste instructional time gathering assessment information when we have enough to serve the purpose. So one key to success (accuracy of assessment and efficiency) is sound sampling.

Another source of mismeasurement can be the exercises, questions, or problems posed on a paper-and-pencil test. Here the pitfalls include such things as unclear instructions, ambiguous wording, and failure to reflect the target in the exercise, such as assessing only recall when the target was higher order thinking. Clear, focused exercises calculated to reflect a precise target prevent problems in assessment.

Mismeasurement can arise from our evaluations of student responses. With objective test formats (right or wrong answers), this is not a problem as long as the scoring key is accurate (a consideration not to be taken lightly). But with subjective ratings, such as in the case of performance assessment, the idiosyncrasies of the rater can become a source of difficulties. To avoid these, clear and explicit evaluation criteria reflecting the desired achievement are a must. Further, careful rater training is the key to the effective application of these criteria. Here's the test of objectivity: If two independent evaluators were to observe the same performance, would they agree on the level of proficiency demonstrated without conversing about it?

Table 2
Matching Assessment Targets and Methods

Target	Method		
	Paper and pencil	Performance assessment	Personal communication
Knowledge	1		2
Thinking	3	4	5
Behavior		6	
Product		7	
Affect	8	9	10

Note. The number entries under "Method" represent the following: 1—Objective test items provide an excellent means to sample mastery of facts and information; 2—Objective questions asked during instruction and answered right or wrong can indicate mastery of knowledge; 3—Essay format works well here, but so do objective formats for many (but not all) valued kinds of thinking; 4—Given complex problems to solve, do students go through the proper steps? 5—Thinking-skills questions, followed up by additional probes, can provide insights into thought processes; 6—What better way to assess speaking (or any other similar skill) than by watching it happen? 7—Products created by students—such as writing samples, for example—contain evidence of the ability to communicate; 8—Questionnaires can serve to provide evidence of attitudes; 9—Effort, for example, is reflected in work-related behaviors; and 10—We can find out how students feel about things by talking to them.

Table 3
Some Practical Considerations in Assessment

Methods	Objective form		Subjective form	
	Strengths	Limits	Strengths	Limits
Paper and pencil	<ul style="list-style-type: none"> Allows for broad sampling, with many items per unit of time Reuse possible Can be scored efficiently 	<ul style="list-style-type: none"> Requires technical knowledge to develop Development takes time Limits targets that can be assessed 	<ul style="list-style-type: none"> Allows assessment of complex thinking Can be developed quickly 	<ul style="list-style-type: none"> Requires expertise to score Takes time to score Limits sample; few items per unit of time
Performance assessment	<ul style="list-style-type: none"> Assesses behavior/product outcomes Permits student to be trained as raters, thus encouraging better performance 	<ul style="list-style-type: none"> Limits targets; few targets translate into checklists Requires rater training, which takes time 	<ul style="list-style-type: none"> Assesses process/product outcomes Permits students to be trained as raters, thus encouraging better performance 	<ul style="list-style-type: none"> Requires time/expertise to develop scoring criteria Requires rater training, which takes time
Personal communication	<ul style="list-style-type: none"> Can assess quickly Can sample group achievement Permits follow-up questions Encourages valuable personal contact 	<ul style="list-style-type: none"> Can be difficult to sample achievement of all individuals Can be difficult to sample broad domains Record keeping can be difficult 	<ul style="list-style-type: none"> Permits follow-up probes Encourages valuable personal contact 	<ul style="list-style-type: none"> Demands time for one-on-one contact Record keeping can be difficult

Mismeasurement can also arise from students themselves. Students can intentionally or unintentionally keep us from learning their true level of achievement. Unintentional sources of inaccurate assessment include reading deficiencies that cause inappropriately low scores on multiple-choice tests of knowledge, writing deficiencies that cause inappropriately low scores in essay tests, or problems in health or well-being that can cost the student his or her concentration at test time. Intentional sources of interference include cheating, shyness in assessments based on public communication, or personal conflicts with teachers that inhibit communication.

The meaning of quality. Teachers, students, and their families count on sound classroom assessments to drive the teaching/learning process. Regardless of their form, classroom assessments are sound only when they (a) arise from and systematically reflect a clear and specific achievement target, (b) sample student performance in a way that represents all key aspects of the target without wasting time with too much assessment, (c) control for all relevant sources of extraneous interference that can cause us to mismeasure, and (d) arise from a clear sense of purpose. Sound classroom assessment training provides assessors with the knowledge and skills they need to meet these standards when using any and all of the various kinds of assessment methods teachers use.

Issues to ponder. Do you know how to link achievement targets to assessments so as to assess accurately? What will it take by way of preparation on your part to assure that students and parents—both critical users of classroom assessment results—clearly understand the meaning of the results of any particular assessment? Confident, competent classroom assessor/teachers artfully orchestrate various targets, assessment methods, and instructional tactics. What professional development will it take on your part to be sure you can blend sound assessment into your teaching/learning environment?

Summary

The purpose of this module is not to prepare competent classroom assessors. Obviously that will take more instruction and professional development than can be provided in these few brief pages. Rather, the module is written to provide teachers and those in positions of instructional leadership with a clear sense of what a competent classroom assessor knows and can do.

Teachers who prepare for assessment with a clear sense of what they wish to assess and why they wish to assess it lay a solid foundation for the selection and use of proper assessment methods. We have many assessment tools at our disposal. They are not interchangeable. Different methods are required to assess different kinds of targets in different contexts. But one thing all methods have in common is that they carry with them specific rules of evidence for their proper use in the classroom. Teachers assess well when they know their purpose, target, and methods—when they know the meaning of assessment quality and apply it day to day in their classrooms.

Self-Test

Read the following scenario and select the assessment method you would use in each context: A science teacher in a large high school teaches four introductory biology classes each day. Each class contains 30 to 35 students. This teacher has prepared a 3-week unit on cells. Each week has a different achievement target. Read each target and context below and determine, for each, if you would select (a) a paper-and-pencil test, (b) a performance assessment (observation and evaluation of achieve-

ment-related behavior of product), or (c) an assessment based on personal communication (e.g., interview or discussion).

1. The valued outcome of week #1 is student mastery of content knowledge about cells. During this week, the teacher wants the students to build a foundation of knowledge before going on to advanced topics. It is Friday of week #1 and the teacher needs to assess mastery of content in all 120–130 students before proceeding into week #2. What assessment method would you recommend and why?
2. During week #2 the target is to have students learn to use the knowledge they mastered to solve certain kinds of practical higher order thinking problems. The teacher wants them to be able to analyze, compare, infer, etc. It is Friday of week #2. What assessment method would you use?
3. The focus of week #3 is different. The teacher divides students into cooperative teams and gives them a laboratory problem to solve. They are to solve it and prepare a presentation to the class on their solution as well as the process they used to arrive at it. The goal is to demonstrate the ability to communicate in science. What assessment method is called for?

Answers to Self-Test

1. Most would choose an objective paper-and-pencil test here. It can (a) accurately reflect the target (mastery of knowledge) and (b) produce needed information very efficiently (critical in this context). But to use this method effectively, what must the teacher possess? The teacher must have a crystal clear vision of the content domain to be mastered and the craft knowledge of paper-and-pencil test development needed to sample that domain with sound test items.
2. You can choose an objective test here too! But to make this work, the teacher must (a) possess a clear vision of the meaning of higher order thinking, (b) have selected kinds of thinking that can be translated into objective test items, and (c) have mastered the craft knowledge needed to write sound test items of this sort. The teacher also could have selected performance assessment here if he or she had mastered the craft knowledge needed to (a) devise strong exercises to sample performance and (b) use labor-intensive methods efficiently, given the large number of students and short time frame in this context.
3. The only choice in this case is observation of and judgment about the presentations—performance assessment. However, to make this work, once again the teacher must (a) possess a specific vision of what it means to communicate effectively in science and (b) have mastered craft knowledge about the design and use of performance assessment. Most importantly, the teacher must know the tricks of the trade when it comes to the efficient use of labor-intensive methods. These can include student peer assessment and self-assessment, group assessment with individual accountability, and the use of various student sampling techniques.

References

- Schafer, W. D., & Lissitz, R. W. (1987). Measurement training for school personnel: Recommendations and reality. *Journal of Teacher Education*, 38(3), 57–63.
- Stiggins, R. J., & Conklin, N. F. (1992). *In teachers' hands: Investigating the practices of classroom assessment*. Albany, NY: SUNY Press.