



## Unlocking the Power of the Teacher-Made Test



***Log On. Let's Talk.***

[www.ets.org/letstalk](http://www.ets.org/letstalk)

*Listening. Learning. Leading.*

## Introduction

Classroom assessment ranks among a teacher's most essential educational tools. Well-constructed teacher-made tests can:

- provide teachers with the means to gather evidence about what their students know and can do
- help instructors identify students' strengths and weaknesses
- keep tabs on student learning and progress
- help teachers plan and conduct future instruction
- motivate and shape learning and instruction
- guide students toward improving their own performances
- gauge whether students are mastering district, state, and national education standards
- determine if students are prepared for the high-stakes state or district tests

By unlocking the power of effective classroom assessment, teachers can accomplish all of the above and more. In the era of accountability and high-stakes decision making, teacher-made tests can no longer be viewed as simply a means to gather grades for the end of the marking period report cards.

## Formative vs. Summative Assessment

In general, all tests can fall under one of two major subheadings: formative assessments or summative assessments.

Formative assessments are those tools teachers use to monitor student performance on an ongoing basis. These can range from something as simple as the daily judgments teachers make about a student's oral response to questions that arise in classroom discussions to more formal paper and pencil tests.

Summative assessments, on the other hand, gather evidence about cumulative student learning at the end of an activity, unit, marking period or school year. These types of assessments would include, among others, the final exam or the chapter test.

Whether they use formative or summative methods, effective teachers constantly monitor and revise instructional plans based on their students' educational progress and needs. Assessment, whether formal or informal, plays a vital part in this ongoing

process, providing the evidence teachers need to determine whether or not their students have achieved the educational goals set out for them.

## Capturing the Evidence

Classroom assessments can be thought of as evidence capturing devices or tools. The evidence the teacher seeks to gather is used to show or prove the students' knowledge and ability. Just like a good detective, the classroom teacher must consider several things in selecting the tools used to gather the evidence:

- What do I think my students should know? What are my expectations for their knowledge base? This could be based on lesson goals and objectives, curriculum or course content goals, district or state standards, etc.
- How would I describe my students after they experience these teaching episodes; how would I capture this description?
- Of all the things I've taught, what are the most important concepts and what should be assessed?
- What is the best tool for capturing this evidence?
- How will I use this evidence in the future? Is it simply to "keep score" or will it cause me to reflect on my teaching and my future planning?
- How will I report this evidence and to whom?

## Elements of an Effective Teacher-Made Test

In reflecting on these questions, teachers begin to realize the power of classroom assessment, and that this tool can be as important in the teaching and learning process as class discussions, small group activities, or any other teaching strategy. To unleash this power, and to ensure that classroom instructional time devoted to the assessment is used wisely, teachers must carefully plan and design the test. A poorly chosen or designed assessment will fail to provide the evidence of student learning, or worse, will provide misleading information.

It is imperative that the teacher employs a systematic process for developing and using the assessment tool. That process should begin with the instructor asking a few basic but essential questions:

- What am I trying to find out about my students' learning? (That is, what student/standards/goals/outcomes am I measuring?)

- What kind of evidence do I need to show that my students have achieved the goals that I'm trying to measure?
- What kind of assessment will give me that evidence?

To respond to these questions, the teacher must consider these elements of the test design process:

- appropriateness
- relevancy
- expectations for learning
- multiple evidence
- planning
- fairness
- assessing the assessment

## Is it Appropriate?

If you were to visit your doctor and he or she used a thermometer to determine your blood pressure you might have cause for concern. Like the classroom test, the medical thermometer is an evidence-gathering device (to determine one's body temperature). Both the physician and the classroom teacher must select the best device based upon the type of evidence it was meant to provide, and not one that provides evidence of something else.

What would happen if a teacher decides to use word problems, rather than number problems, to determine whether third-graders know their multiplication facts? One outcome that we might imagine is that the teacher could not be certain if the test was measuring math ability or reading skill. Having to read and understand the questions could get in the way of a student being able to demonstrate that he or she can, for instance, multiply  $8 \times 6$  and come up with 48. Of course, if the teacher wants to know whether students can apply their multiplication skills to realistic situations, the word problem task would be appropriate and fair.

At the same time, the assessment must measure the knowledge, skills, and/or abilities the teacher feels are important and do this in an appropriate way. If the goal is to test for retention of facts, then a cut-and-dried factual test (e.g., multiple-choice or fill-in-the-blank) may be the best assessment choice. On the other hand, measuring students' conceptual understanding, ability to analyze data, ability to perform tasks, or their collaborative skills would probably require more complex forms of assessment.

## Is it Relevant?

An assessment task should make sense in terms of the assessment situation as well as the type of knowledge or skill that's being assessed. It should also provide relevant information based on what students should have learned in class. For example, it would be inappropriate to ask ninth-grade, French-class students to carry on a conversation in French about the income tax system, if it has nothing to do with what they had learned in French class, and since it probably has no relevance in their lives. But asking them to carry on a discussion in French about a class trip would be very appropriate.

Ideally, an assessment should also reflect real-world applications of knowledge and understanding. Although developing such assessments is not always practical, assessments based on situations relevant to students' own world experiences can motivate them to put forth their best performances. If they don't understand why they may need to know something, they won't be as likely to do so.

## What are the Expectations for Learning?

Before administering an assessment designed to measure what students have learned in class, teachers need to ask themselves: "Based on what I've taught in class, can my students be expected to answer this?" If correctly completing the assessment requires knowledge or skills that have not been emphasized in class, or that the students are unlikely to have mastered, the assessment will not provide an accurate or fair evaluation of whether the students have learned the material. Of course, if the goal is to find out what students already know or understand before a unit of instruction, then a well-thought-out assessment can provide useful information for planning future lessons. Of course, informing future lesson planning should be one of the key uses of any assessment.

## One Test or Multiple Sources of Evidence?

While a single clue at a crime scene might allow a detective to make some assumptions, the judge and jury will probably need a lot more evidence before making a decision about guilt or innocence. As a result, experienced and competent investigators will employ many tools to gather multiple types of evidence from sources such as fingerprints, DNA samples, other physical clues, as well as actual interviews of witnesses and suspects. Similarly, teachers

should rely on all sorts of options available to help them gather evidence of their students' learning. These range from the informal, day-to-day ways teachers size up their students' progress, such as observation and questioning strategies, to traditional paper-and-pencil tests (multiple-choice and short-answer ones, for example), to more elaborate forms of assessment, such as essays or problem solving activities. Just as a detective must gather many sources of evidence to build a convincing case, so must a teacher use many sources of evidence to accurately interpret what each student really knows and can do.

Limiting themselves to using only one or two assessment methods, no matter how reliable or valuable, limits teachers' ability to fully understand the range of their students' knowledge and skills. Conversely, providing students with different kinds of opportunities to show what they know gives teachers a broader, better understanding of each student's talents and abilities.

## Planning the Assessments

Teachers should use the following guidelines when planning any assessment:

- Have the purpose of the test clearly in mind.
- Determine what type of assessment will be most appropriate for the situation, based on the nature of what you are teaching, the purpose of the instruction, and what you want to find out.
- If the purpose of the assessment is to determine how well students have mastered a particular unit of study, make sure the test parallels the work covered in class. And, to be able to discriminate among levels of learning, avoid making the assessment overly difficult or easy.
- If the assessment is a selected-response or fill-in-the-blank test that will be used to diagnose basic skills, it should contain at least 10 questions — preferably more — for each skill area. The questions pertaining to each skill area should be considered a subtest, and these subtests should yield separate scores on the various elements needed for mastery of the skill.
- If the major purpose of the test is to rank a selected group of students in order of their achievement, the questions should cover critical points of learning. Questions on critical points often require understanding implications, applying information, and reorganizing data.

The questions should challenge students to do more than memorize and recall facts.

- Focus on assessing the most important and meaningful information, rather than small, irrelevant facts. For example, rather than asking, "How many vitamins are essential for humans? A.7 B.13 C.15 D.23," consider asking "Name at least seven vitamins that are essential for humans and explain why they are essential."
- Never use questions or inconsequential details just to trick students.
- Create a test blueprint that will clearly describe the important content areas to be tested, the number and type of items that will get at each content area, the scoring value for the items, the length of time for the test administration, and other critical test components.

## Is it Fair?

To be valid, classroom assessments need to be fair. In assessment terms, that means all students must be given an equal chance to show what they know and can do. An assessment is not fair if it:

- measures things unrelated to its objectives is biased

Tests should be designed so that they are focused on the instruction that preceded the assessment, and that ensure that testing, teaching and curriculum are all tightly aligned.

"Bias" is said to exist if the assessment includes content that offends or unfairly penalizes test takers because of personal characteristics such as gender, ethnicity, socio-economic status, religion, or sexual orientation. Assessment bias affects validity because it may negatively influence students' attitudes toward, and performance on, the assessment. For example, an assessment that includes language or content that offends a specific group of students may hurt the performance of those students by causing them to focus more on the offensive language than on performing at their best. Test takers may also be disadvantaged if the assessment includes content that, although not offensive, assumes prior knowledge likely to be unfamiliar to one particular group of test takers but familiar to another. An obvious example is the need to know something that can only be gained by visiting an art museum, if some groups of students have no personal experience of ever visiting a museum with family or friends.

Tests should be designed to afford students multiple opportunities to tell what they know about a particular subject, not to present them with difficult, if not impossible, tasks. One way to let students shine is to include a bonus question at the end of the test that asks something like, “Take this opportunity to tell me something about this topic that was not included on the test.”

## Assess the Assessment

In assessment, wording is critical. Unclear directions can confuse test takers and negatively affect their responses, which can lead to inaccurate, and therefore useless, information about what the students actually know and can do. Wording in multiple-choice type items is especially important. Being precise in the question and in the options or choices prevents misunderstanding and provides more reliable evidence of what students know.

After an assessment has been administered, teachers can ask students how they interpreted the questions, particularly if the questions elicited unexpected results. When assessments give unexpected results — for example, the entire class bombs an assessment, or the students’ responses are not consistent with the type of work the teacher was looking for — it’s important to take a good hard look at both the assessment and the way it was administered to determine whether it was flawed in some way.

- Did students who are more able, based on other evidence, do well on the assessment? If not, something might be wrong with it. You might consider analyzing the questions or tasks to make sure each is accurate, valid, fair, and reliable. On the other hand, if the assessment presents a type of task that your students might not be familiar with (e.g., a complex, nonroutine type of problem), students who performed poorly may simply have had difficulty with that particular type of assessment.
- Did students answer the assessment appropriately but not give the answers you were looking for? Then check to see if the task was well-defined and clearly written. Students can’t be expected to give adequate responses if they aren’t sure what kind of response is expected of them.
- If the entire class failed the test, it might indicate that the material wasn’t taught adequately, or the assessment was so poorly written that the students were unable to apply their knowledge appropriately.

- Other things to consider when poor results are obtained are external, test administration issues [e.g., uncomfortable room temperature, administration right after a long weekend, external noise distractions, unsettling school or community news].

One of the most effective ways to improve a classroom assessment is to review it before administering it. If possible, wait at least one day after the assessment was written before performing the review. Then ask a colleague to review the assessment. As part of this review process, have someone who did not write the task (a colleague or even a family member) solve the task.

## During the review, check to see that:

- directions are clear
- content is accurate
- questions or tasks are representative of the topics or skills emphasized during instruction; knowledge or skills that were not covered in class are not being unintentionally evaluated
- the type of assessment used is compatible with the method of instruction used in the classroom and the standard being measured
- the assessment will contribute to the instructor’s understanding of what the students know and can do
- the assessment can be completed in the allotted time
- the assessment is fair; all instances of offensive language, elitism, and bias have been eliminated

When teachers begin to analyze assessment results, they should look for two things:

1. Does the question or task provide accurate information?

Did all the students do poorly on the same question or set of questions? Maybe certain questions are confusing or misleading, or perhaps the concept is simply not yet well understood by the students and should be retaught. Having students explain why they answered a question in a certain way can be very enlightening to the teacher about whether the problem is in the question (or task), or in students’ understanding of the concept being assessed.

2. Each student's strengths and weaknesses: are they based on his or her patterns or performance?

This information can help teachers tailor the next round of instruction to either remedy problems or build on strength. For example, if a particular group of students has difficulty with one set of items that measures a similar set of skills, these students might need extra instruction or a different kind of instruction. Or, if everybody in the class had difficulty with a particular issue that the teacher thought was emphasized in class, then the teacher needs to determine if there was a problem with the instruction and/or material.

By taking the time to create fair, focused, and well-thought-out assessments, teachers can have confidence in the evidence gathered and make meaningful judgments about student performance and future instructional plans and decisions.

This article was based on the ETS Focus publication *Letting Students Shine: Assessment to Promote Student Learning*. Written by Amada McBride, 1999.

