

Authentic Assessment Toolbox

Created by Jon Mueller

<http://jonathan.mueller.faculty.noctrl.edu/toolbox/whatisit.htm#names>

What is Authentic Assessment?

► [Definitions](#)

► [What Does Authentic Assessment Look Like?](#)

► [How is Authentic Assessment Similar to/Different from Traditional Assessment?](#)

- [Traditional Assessment](#)
- [Authentic Assessment](#)
- [Authentic Assessment Complements Traditional Assessment](#)
- [Defining Attributes of Authentic and Traditional Assessment](#)
- [Teaching to the Test](#)

► [Alternative Names for Authentic Assessment](#)

Definitions

A form of assessment in which students are asked to perform real-world tasks that demonstrate meaningful application of essential knowledge and skills -- Jon Mueller

"...Engaging and worthy problems or questions of importance, in which students must use knowledge to fashion performances effectively and creatively. The tasks are either replicas of or analogous to the kinds of problems faced by adult citizens and consumers or professionals in the field." -- Grant Wiggins -- ([Wiggins, 1993, p. 229](#)).

"Performance assessments call upon the examinee to demonstrate specific skills and competencies, that is, to apply the skills and knowledge they have mastered." -- Richard J. Stiggins -- ([Stiggins, 1987, p. 34](#)).

What does Authentic Assessment look like?

An authentic assessment usually includes a task for students to perform and a rubric by which their performance on the task will be evaluated. Click the following links to see many examples of authentic tasks and rubrics.

- [Examples](#) from teachers in my Authentic Assessment course

How is Authentic Assessment similar to/different from Traditional Assessment?

The following comparison is somewhat simplistic, but I hope it illuminates the different assumptions of the two approaches to assessment.

Traditional Assessment

By "traditional assessment" (TA) I am referring to the forced-choice measures of multiple-choice tests, fill-in-the-blanks, true-false, matching and the like that have been and remain so common in education. Students typically select an answer or recall information to complete the assessment. These tests may be standardized or teacher-created. They may be administered locally or statewide, or internationally.

Behind traditional and authentic assessments is a belief that the primary mission of schools is to help develop productive citizens. That is the essence of most mission statements I have read. From this common beginning, the two perspectives on assessment diverge. Essentially, TA is grounded in educational philosophy that adopts the following reasoning and practice:

1. A school's mission is to develop productive citizens.
2. To be a productive citizen an individual must possess a certain body of knowledge and skills.
3. Therefore, schools must teach this body of knowledge and skills.
4. To determine if it is successful, the school must then test students to see if they acquired the knowledge and skills.

In the TA model, the curriculum drives assessment. "The" body of knowledge is determined first. That knowledge becomes the curriculum that is delivered. Subsequently, the assessments are developed and administered to determine if acquisition of the curriculum occurred.

Authentic Assessment

In contrast, authentic assessment (AA) springs from the following reasoning and practice:

1. A school's mission is to develop productive citizens.
2. To be a productive citizen, an individual must be capable of performing meaningful tasks in the real world.
3. Therefore, schools must help students become proficient at performing the tasks they will encounter when they graduate.
4. To determine if it is successful, the school must then ask students to perform meaningful tasks that replicate real world challenges to see if students are capable of doing so.

Thus, in AA, assessment drives the curriculum. That is, teachers first determine the tasks that students will perform to demonstrate their mastery, and then a curriculum is developed that will enable students to perform those tasks well, which would include the acquisition of essential knowledge and skills. This has been referred to as *planning backwards* (e.g., [McDonald, 1992](#)).

If I were a golf instructor and I taught the skills required to perform well, I would not assess my students' performance by giving them a multiple choice test. I would put them out on the golf course and ask them to perform. Although this is obvious with athletic skills, it is also true for academic subjects. We can teach students how to *do* math, *do* history and *do* science, not just *know* them. Then, to assess what our students had learned, we can ask students to perform tasks that "replicate the challenges" faced by those using mathematics, doing history or conducting scientific investigation.

Authentic Assessment Complements Traditional Assessment

But a teacher does not have to choose between AA and TA. It is likely that some mix of the two will best meet your needs. To use a silly example, if I had to choose a chauffeur from between someone who passed the *driving* portion of the driver's license test but failed the *written* portion or someone who failed the driving portion and passed the written portion, I would choose the driver who most directly demonstrated the ability to drive, that is, the one who passed the driving portion of the test. However, I would *prefer* a driver who passed both portions. I would feel more comfortable knowing that my chauffeur had a good knowledge base about driving (which might best be assessed in a traditional manner) and was able to apply that knowledge in a real context (which could be demonstrated through an authentic assessment).

Defining Attributes of Traditional and Authentic Assessment

Another way that AA is commonly distinguished from TA is in terms of its defining attributes. Of course, TA's as well as AA's vary considerably in the forms they take. But, typically, along the continuums of attributes listed below, TA's fall more towards the left end of each continuum and AA's fall more towards the right end.

Traditional ----- Authentic

Selecting a Response ----- Performing a Task

Contrived ----- Real-life

Recall/Recognition ----- Construction/Application

Teacher-structured ----- Student-structured

Indirect Evidence ----- Direct Evidence

Let me clarify the attributes by elaborating on each in the context of traditional and authentic assessments:

Selecting a Response to Performing a Task: On traditional assessments, students are typically given several choices (e.g., a,b,c or d; true or false; which of these match with those) and asked to select the right answer. In contrast, authentic assessments ask students to demonstrate understanding by performing a more complex task usually representative of more meaningful application.

Contrived to Real-life: It is not very often in life outside of school that we are asked to select from four alternatives to indicate our proficiency at something. Tests offer these contrived means of assessment to increase the number of times you can be asked to demonstrate proficiency in a short period of time. More commonly in life, as in authentic assessments, we are asked to demonstrate proficiency by doing something.

Recall/Recognition of Knowledge to Construction/Application of Knowledge: Well-designed traditional assessments (i.e., tests and quizzes) can effectively determine whether or not students have acquired a body of knowledge. Thus, as mentioned above, tests can serve as a nice complement to authentic assessments in a teacher's assessment portfolio. Furthermore, we *are* often asked to recall or recognize facts and ideas and propositions in life, so tests are somewhat authentic in that sense. However, the demonstration of recall and recognition on tests is typically much less revealing about what we really know and can do than when we are asked to construct a product or performance out of facts, ideas and propositions. Authentic assessments often ask students to analyze, synthesize and apply what they have learned in a substantial manner, and students create new meaning in the process as well.

Teacher-structured to Student-structured: When completing a traditional assessment, what a student can and will demonstrate has been carefully structured by the person(s) who developed the test. A student's attention will understandably be focused on and limited to what is on the test. In contrast, authentic assessments allow more student choice and construction in determining what is presented as evidence of proficiency. Even when students cannot choose their own topics or formats, there are usually multiple acceptable routes towards constructing a product or performance. Obviously, assessments more carefully controlled by the teachers offer advantages and disadvantages. Similarly, more student-structured tasks have strengths and weaknesses that must be considered when choosing and designing an assessment.

Indirect Evidence to Direct Evidence: Even if a multiple-choice question asks a student to analyze or apply facts to a new situation rather than just recall the facts, and the student selects the correct answer, what do you now know about that student? Did that student get lucky and pick the right answer? What thinking led the student to pick that answer? We really do not know. At best, we can make some inferences about what that student might know and might be able to do with that knowledge. The evidence is very indirect, particularly for claims of meaningful application in complex, real-world situations. Authentic assessments, on the other hand, offer more direct evidence of application and construction of knowledge. As in the golf example above, putting a golf student on the golf course to play provides much more direct evidence of proficiency than giving the student a written test. Can a student effectively critique the arguments someone else has presented (an important skill often required in the real world)? Asking a student to write a critique should provide more direct evidence of that skill than asking the student a series of multiple-choice, analytical questions about a passage, although both assessments may be useful.

Teaching to the Test

These two different approaches to assessment also offer different advice about teaching to the test. Under the TA model, teachers have been discouraged from teaching to the test. That is because a test usually assesses a sample of students' knowledge and understanding and assumes that students' performance on the sample is representative of their knowledge of all the relevant material. If teachers focus primarily on the sample to be tested during instruction, then good performance on that sample does not necessarily reflect knowledge of all the material. So, teachers hide the test so that the sample is not known beforehand, and teachers are admonished not to teach to the test.

With AA, teachers are *encouraged* to teach to the test. Students need to learn how to perform well on meaningful tasks. To aid students in that process, it is helpful to show them models of good (and not so good) performance. Furthermore, the student benefits from seeing the task rubric ahead of time as well. Is this "cheating"? Will students then just be able

to mimic the work of others without truly understanding what they are doing? Authentic assessments typically do not lend themselves to mimicry. There is not one correct answer to copy. So, by knowing what good performance looks like, and by knowing what specific characteristics make up good performance, students can better develop the skills and understanding necessary to perform well on these tasks. (For further discussion of teaching to the test, see [Bushweller](#).)

Alternative Names for Authentic Assessment

You can also learn something about what AA is by looking at the other common names for this form of assessment. For example, AA is sometimes referred to as

- **Performance Assessment** (or Performance-based) -- so-called because students are asked to *perform* meaningful tasks. This is the other most common term for this type of assessment. Some educators distinguish performance assessment from AA by defining performance assessment as performance-based as Stiggins has above but with no reference to the *authentic* nature of the task (e.g., [Meyer, 1992](#)). For these educators, authentic assessments are performance assessments using real-world or authentic tasks or contexts. Since we should not typically ask students to perform work that is not authentic in nature, I choose to treat these two terms synonymously.
- **Alternative Assessment** -- so-called because AA is an *alternative* to traditional assessments.
- **Direct Assessment** -- so-called because AA provides more *direct* evidence of meaningful application of knowledge and skills. If a student does well on a multiple-choice test we might infer *indirectly* that the student could apply that knowledge in real-world contexts, but we would be more comfortable making that inference from a direct demonstration of that application such as in the golfing example above.

Why Use Authentic Assessment?

The question "Why use authentic assessment?" is not meant to suggest that you have to choose between traditional assessments such as tests and more authentic or performance assessments. Often, teachers use a mix of traditional and authentic assessments to serve different purposes. This section, then, attempts to explain why teachers might choose authentic assessments for certain types of judgments and why authentic assessments have become more popular in recent years.

Authentic Assessments are Direct Measures

We do not just want students to *know* the content of the disciplines when they graduate. We, of course, want them to be able to *use* the acquired knowledge and skills in the real world. So, our assessments have to also tell us if students can apply what they have learned in authentic situations. If a student does well on a test of knowledge we might infer that the student could also apply that knowledge. But that is rather indirect evidence. I could more directly check for the ability to apply by asking the student to use what they have learned in some meaningful way. To return to an example I have used elsewhere, if I taught someone to play golf I would not check what they have learned with just a written test. I would want to see more direct, authentic evidence. I would put my student out on a golf course to play. Similarly, if we want to know if our students can interpret literature, calculate potential savings on sale items, test a hypothesis, develop a fitness plan, converse in a foreign language, or apply other knowledge and skills they have learned, then authentic assessments will provide the most direct evidence.

Can you think of professions which require some direct demonstration of relevant skills before someone can be employed in that field? Doctors, electricians, teachers, actors and others must all provide direct evidence of competence to be hired. Completing a written or oral test or interview is usually not sufficient. Shouldn't we ask the same of our students before we say they are ready to graduate? Or pass a course? Or move on to the next grade?

Authentic Assessments Capture Constructive Nature of Learning

A considerable body of research on learning has found that we cannot simply be fed knowledge. We need to construct our own meaning of the world, using information we have gathered and were taught and our own experiences with the world (e.g., [Bransford & Vye, 1989](#); [Forman & Kuschner, 1977](#); [Neisser, 1967](#); [Steffe & Gale, 1995](#); [Wittrock, 1991](#)). Thus, assessments cannot just ask students to repeat back information they have received. Students must also be asked to demonstrate that they have accurately constructed meaning about what they have been taught. Furthermore, students must be given the opportunity to engage in the construction of meaning. Authentic tasks not only serve as assessments but also as vehicles for such learning.

Authentic Assessments Integrate Teaching, Learning and Assessment

Authentic assessment, in contrast to more traditional assessment, encourages the integration of teaching, learning and assessing. In the "traditional assessment" model, teaching and learning are often separated from assessment, i.e., a test is administered after knowledge or skills have (hopefully) been acquired. In the authentic assessment model, the same authentic task used to measure the students' ability to apply the knowledge or skills is used as a vehicle for student learning. For example, when presented with a real-world problem to solve, students are learning in the process of developing a solution, teachers are facilitating the process, and the students' solutions to the problem becomes an assessment of how well the students can meaningfully apply the concepts.

Authentic Assessments Provide Multiple Paths to Demonstration

We all have different strengths and weaknesses in how we learn. Similarly, we are different in how we can best *demonstrate* what we have learned. Regarding the traditional assessment model, answering multiple-choice questions does not allow for much variability in how students demonstrate the knowledge and skills they have acquired. On the one hand, that is a strength of tests because it makes sure everyone is being compared on the same domains in the same manner which increases the consistency and comparability of the measure. On the other hand, testing favors those who are better test-takers and does not give students any choice in how they believe they can best demonstrate what they have learned.

Thus, it is recommended (e.g., [Wiggins, 1998](#)) that multiple and varied assessments be used so that 1) a sufficient number of samples are obtained (multiple), and 2) a sufficient variety of measures are used (varied). Variety of measurement can be accomplished by assessing the students through different measures that allows you to see them apply what they have learned in different ways and from different perspectives. Typically, you will be more confident in the students' grasp of the material if they can do so. But some variety of assessment can also be accomplished *within* a single measure. Authentic tasks tend to give the students more freedom in how they will demonstrate what they have learned. By carefully identifying the criteria of good performance on the authentic task ahead of time, the teacher can still make comparable judgments of student performance even though student performance might be expressed quite differently from student to student. For example, the products students create to demonstrate authentic learning on the same task might take different forms (e.g., posters, oral presentations, videos, websites). Or, even though students might be required to produce the same authentic product, there can be room within the product for different modes of expression. For example, writing a good persuasive essay requires a common set of skills from students, but there is still room for variation in how that essay is constructed.

How Do You Create Authentic Assessments?

Authentic Assessment: *Students are asked to perform real-world tasks that demonstrate meaningful application of essential knowledge and skills*

Fortunately, you do not have to develop an authentic assessment from scratch. You may already be using [authentic tasks](#) in your classroom. Or, you may already have the [standards](#) written, the first and most important step in the process. Perhaps you have a task but need to more clearly articulate the [criteria](#) for evaluating student performance on the task. Or, you may just want to develop a [rubric](#) for the task. Wherever you are in the process, you can use the information on this page (and the ones that follow it) to help you through the steps of creating authentic assessments. If at any time the terminology is confusing, click a link to that concept or go to the [glossary](#).

I tend to think of authentic assessment development in terms of four questions to be asked. Those questions are captured in the following graphic:

Questions to Ask:

1) What should students know and be able to do?
This list of knowledge and skills becomes your . . .

STANDARDS



2) What indicates students have met these standards?
To determine if students have met these standards, you
will design or select relevant . . .

AUTHENTIC TASKS



3) What does good performance on this task look like?
To determine if students have performed well on the task,
you will identify and look for characteristics of good
performance called . . .

CRITERIA



4) How well did the students perform?
To discriminate among student performance
across criteria, you will create a . . .

RUBRIC



5) How well should most students perform?
The minimum level at which you would want most students to perform
is your ...



CUT SCORE or BENCHMARK

6) What do students need to improve upon?
Information from the rubric will give students feedback and allow you
to ...



ADJUST INSTRUCTION

Summary of Steps

1. Identify your **standards** for your students.
2. For a particular standard or set of standards, develop a **task** your students could perform that would indicate that they have met these standards.
3. Identify the characteristics of good performance on that task, the **criteria**, that, if present in your students' work, will indicate that they have performed well on the task, i.e., they have met the standards.
4. For each criterion, identify two or more levels of performance along which students can perform which will sufficiently discriminate among student performance for that criterion. The combination of the criteria and the levels of performance for each criterion will be your **rubric** for that task (assessment).

Now, I will guide you through each these four steps for creating an authentic assessment in more detail.

Step 1: Identify the Standards

For any type of assessment, you first must know where you want to end up. What are your goals for your students? An assessment cannot produce *valid* inferences unless it measures what it is intended to measure. And it cannot measure what it is intended to measure unless the goal(s) has been clearly identified. So, completing the rest of the following steps will be unproductive without clear goals for student learning.

Standards, like goals, are statements of what students should know and be able to do. However, standards are typically more narrow in scope and more amenable to assessment than goals. (Before going further, I would recommend that you read the section on [Standards](#) for a fuller description of standards and how they are different from goals and objectives.)

- ▶ [What Do Standards Look Like?](#)
- ▶ [How do you get Started Writing Standards?](#)
- ▶ [What are Some Guidelines to Follow in Developing Standards?](#)
- ▶ [Workshop: Writing a Good Standard](#)

What Do Standards Look Like?

Standards are typically one-sentence statements of what students should know and be able to do at a certain point. Often a standard will begin with a phrase such as "Students will be able to ..." (SWBAT). For example,

Students will be able to add two-digit numbers.

Or, it might be phrased

Students will add two-digit numbers.

A student will add two-digit numbers.

Or just

Identify the causes and consequences of the Revolutionary War.

Explain the process of photosynthesis.

More examples:

[Extensive set of links](#) to standards organized by subject and state (originally from the Putnam Valley School District)

[Standards examples \(this site\)](#)

Also, read the section on [types of standards](#) to see how standards can address course content, or process skills or attitudes towards learning.

How Do You Get Started?

I recommend a three-step process for writing standards:

1. REFLECT

2. REVIEW

3. WRITE

1. REFLECT

As I will discuss below, there are many sources you can turn to to find examples of goals and standards that might be appropriate for your students. There are national and state standards as well as numerous websites such as those above with many good choices. It is unnecessary to start from scratch. However, before you look at the work of others, which can confine your thinking, I would highly recommend that you, as a teacher or school or district, take some time to examine (or REFLECT upon) what *you* value. What do you really want your students to know and be able to do when they leave your grade or school?

Here is a sample of questions you might ask yourself:

- What do you want students to come away with from an education at _____?
- What should *citizens* know and be able to do?
- If you are writing standards for a particular discipline, what should *citizens* know and be able to do related to your discipline?
- What goals and standards do you share with other disciplines?
- What college preparation should you provide?
- Think of a graduate or current student that particularly exemplifies the set of knowledge and skills that will make/has made that student successful in the real world. What knowledge and skills (related and unrelated to your discipline) does that person possess?
- Ask yourself, "above all else, we want to graduate students who can/will?"
- When you find yourself complaining about what students can't or don't do, what do you most often identify?

As a result of this reflection, you might reach consensus on a few things you most value and agree should be included in the standards. You might actually write a few standards. Or, you might produce a long list of possible candidates for standards. I do not believe there is a particular product you need to generate as a result of the reflection phase. Rather, you should move on to Step 2 (Review) when you are clear about what is most important for your students to learn. For example, reflection and conversation with many of the stakeholders for education led the Maryland State Department of Education to identify the **Skills for Success** it believes are essential for today's citizens. Along with content standards, the high school assessment program in Maryland will evaluate how well students have acquired the ability to learn, think, communicate, use technology and work with others.

2. REVIEW

Did you wake up this morning thinking, "Hey, I'm going to reinvent the wheel today"? No need. There are many, many good models of learning goals and standards available to you. So, before you start putting yours down on paper, REVIEW what others have developed. For example, you can

Look at

- your state goals and standards
- relevant national goals and standards
- other state and local standards already created
 - check out the site mentioned above - Putnam Valley
- your existing goals and standards if you have any
- other sources that may be relevant (e.g., what employers want, what colleges want)

Look for

- descriptions and language that capture what you said you value in Step 1 (REFLECT)
- knowledge and skills not captured in the first step -- should they be included?
- ways to organize and connect the important knowledge and skills

Look to

- develop a good sense of the whole picture of what you want your students to know and to do
- identify for which checkpoints (grades) you want to write standards

3. WRITE

The biggest problem I have observed in standards writing among the schools and districts I have worked with is the missing of the forest for the trees. As with many tasks, too often we get bogged down in the details and lose track of the big picture. I cannot emphasize enough how important it is to periodically step back and reflect upon the process. As you write your standards, ask yourself and your colleagues **guiding questions** such as

- So, tell me again, why do we think this is important?
- Realistically, are they ever going to have to know this/do this/use this?
- How does this knowledge/skill relate to this standard over here?
- We don't have a standard about X; is this really more important than X?
- Can we really assess this? Should we assess it?
- Is this knowledge or skill essential for becoming a productive citizen? How? Why?
- Is this knowledge or skill essential for college preparation?

Yes, you may annoy your colleagues with these questions (particularly if you ask them repeatedly as I would advocate), but you will end up with a better set of standards that will last longer and provide a stronger foundation for the steps that follow in the creation of performance assessments.

Having said that, let's get down to the details. I will offer suggestions for writing specific standards by a) listing some common guidelines for good standards and b) modeling the development of a couple standards much as I would if I were working one-on-one with an educator.

Guidelines for Writing Standards

GUIDELINE #1: For a standard to be amenable to assessment, it must be observable and measurable. For example, a standard such as

"Students *will correctly add* two-digit numbers"

is observable and measurable. However, a standard such as

"Students *will understand how to add* two-digit numbers"

is not observable and measurable. You cannot observe understanding directly, but you can observe performance. Thus, standards should include a verb phrase that captures the direct demonstration of what students know and are able to do.

Some bad examples:

Students *will develop* their persuasive writing skills.

Students *will gain an understanding of* pinhole cameras.

Rewritten as good examples:

Students *will write* an effective persuasive essay.

Students *will use* pinhole cameras to create paper positives and negatives.

GUIDELINE #2: A standard is typically more narrow than a goal and broader than an objective. (See the section on [Standards](#) for a fuller discussion of this distinction.)

Too Broad

Of course, the line between goals and standards and objectives will be fuzzy. There is no easy way to tell where one begins and another one ends. Similarly, some standards will be broader than others. But, generally, a standard is written too broadly if

- it cannot be reasonably assessed with just one or two assessments
- (for content standards) it covers at least half the subject matter of a course or a semester

For example, the [Illinois Learning Standards](#) for [social science](#) lists "Understand political systems, with an emphasis on the United States" as a goal. That is a goal addressed throughout an entire course, semester or multiple courses. The goal is broken down into six standards including "Understand election processes and responsibilities of citizens." That standard describes what might typically be taught in one section of a course or one unit. Furthermore, I feel I could adequately capture a student's understanding and application of that standard in one or two assessments. However, I do not believe I could get a full and rich sense of a student's grasp of the entire goal without a greater number and variety of classroom measures. On the other hand, the standard, "understand election processes and responsibilities of citizens," would not typically be taught in just one or two lessons, so it is broader than an objective. Hence, it best fits the category of a *standard* as that term is commonly used.

Another tendency to avoid that can inflate the breadth of a standard and make it more difficult to assess is the coupling of two or more standards in a single statement. This most commonly occurs with the simple use of the conjunction "and." For example, a statement might read

Students will compare and contrast world political systems **and** analyze the relationships and tensions between different countries.

Although these two competencies are related, each one stands alone as a distinct standard. Additionally, a standard should be assessable by one or two measures. Do I always want to assess these abilities together? I could, but it restricts my options and may not always be appropriate. It would be better to create two standards.

Students will compare and contrast world political systems.

Students will analyze the relationships and tensions between different countries.

In contrast, the use of "and" might be more appropriate in the following standard:

Students will find *and* evaluate information relevant to the topic.

In this case, the two skills are closely related, often intertwined and often assessed together.

Too Narrow

A possible *objective* falling under the social science standard mentioned above that a lesson or two might be built around would be "students will be able to describe the evolution of the voter registration process in this country." This statement would typically be too narrow for a standard because, again, it addresses a relatively small portion of the content of election processes and citizen responsibilities, and because it could be meaningfully assessed in one essay question on a test. Of course, you might give the topic more attention in your government course, so what becomes an objective versus a standard can vary. Also, it is important to note that standards written for larger entities such as states or districts tend to be broader in nature than standards written by individual teachers for their classrooms. A U.S. government teacher might identify 5-15 essential ideas and skills for his/her course and voter registration might be one of them.

As you can see, each of these distinctions and labels are judgment calls. It is more important that you apply the labels *consistently* than that you use a specific label.

Note: You may have noticed that the Illinois Learning Standard that I have been using as an example violates Guideline #1 above -- it uses the verb *understand* instead of something observable. The Illinois Standards avoids this "problem" in

most cases. However, the State addresses it more directly by writing its "benchmark standards" in more observable language. For example, under the general standard "understand election processes and responsibilities of citizens" it states that by early high school (a benchmark) students will be able to "describe the meaning of participatory citizenship (e.g., volunteerism, voting) at all levels of government and society in the United States."

GUIDELINE #3: A standard should not include mention of the specific task by which students will demonstrate what they know or are able to do.

For example, in a foreign language course students might be asked to

Identify cultural differences and similarities between the student's own culture and the target culture using a Venn diagram.

The statement should have left off the last phrase "using a Venn diagram." Completing a Venn diagram is the task the teacher will use to identify if students meet the standard. *How* the student demonstrates understanding or application should not be included with what is to be understood or applied. By including the task description in the standard, the educator is restricted to only using that task to measure the standard because that is what the standard requires. But there are obviously other means of assessing the student's ability to compare and contrast cultural features. So, separate the description of the task from the statement of what the student should know or be able to do; do not include a task in a standard.

GUIDELINE #4: Standards should be written clearly.

GUIDELINE #5: Standards should be written in language that students and parents can understand.

Share your expectations with all constituencies. Students, parents and the community will feel more involved in the process of education. Standards are not typically written in language that early elementary students can always understand, but the standards (your expectations) can be explained to them.

Workshop: Writing a Good Standard

In the "workshops" sprinkled throughout this website I will attempt to capture (and model) the process I follow when assisting someone or some group in developing standards or authentic tasks or rubrics. For this workshop, I will begin with an initial draft of a standard and work with an imaginary educator towards a final product. You can "play along at home" by imagining how you would respond to the educator or to me.

Somewhere in the Smoky Mountains (hey, it's my workshop; I'll host it where I like!)

Educator: How is this for a standard:

I will teach my students what the main themes of Romeo and Juliet are.

Me: First, standards describe what students should know and do, not what the teacher will do. So, standards typically begin, "Students will"

Educator: So, I could change it to

Students will know the main themes of Romeo and Juliet.

Me: Yes, that would be a more appropriate way to begin your standard. Standards also should describe observable and measurable behavior on the student's part so that we can assess it. "Knowing" is not something you can directly observe. So, ask yourself "how could they show me they know?"

Educator: Well, I could have them write a paper explaining the main themes. Maybe I could write a standard saying

Students will write a paper explaining the main themes of Romeo and Juliet.

Me: Can you observe "explaining"?

Educator: Yes, I think so.

Me: Yes, so that verb is a good one for a standard. Are there other ways a student could explain the themes to you besides in a paper?

Educator: Sure. They could do it in a speech, or a poster or on an exam.

Me: Good. You don't want to limit yourself in how you might assess this understanding. So, you usually want to avoid including an assignment or task in your standard. Otherwise, you always have to assign a paper to meet that standard.

Educator: I could say

Students will explain the themes of Romeo and Juliet.

Me: Yes, that is observable and clear. It effectively describes the student learning you said you wanted at the beginning. But let's go back to the main question. You always want to ask yourself "why would I want my students to meet this standard?" Why do you want them to be able to explain the themes of Romeo and Juliet?

Educator: Well, I want my students to be able to pick up a piece of literature and be able to tell what the author's main ideas are, and to find some meaning in it for them.

Me: So, you would like them to do that for literature other than Romeo and Juliet as well?

Educator: Yes, we just always teach Romeo and Juliet.

Me: So, you want to identify what really matters to you, what you really want the students to come away with. Typically, that will go beyond one piece of literature or one author. So, you want to write a standard more generically so that you can choose from a variety of literature and still develop the same knowledge and skills in your students.

Educator: I see. That makes sense. I could say

Students will be able to identify themes across a variety of literature.

Me: Very good. But now I am going to be tough on you. I imagine there are some fourth grade teachers who would tell me they have that same standard for their readers. Is the skill of "identifying a theme" really something your ninth and tenth grade students are learning in your classes or do they come to you with that ability?

Educator: Well, they should have it when they get to me, but many of them still can't identify a theme very well. And, now I am asking them to do it with a more sophisticated piece of literature than fourth graders read.

Me: So, it is certainly appropriate that your students continue to review and develop that skill. But would you hope that your students understanding of theme goes beyond simply being able to identify it in a piece?

Educator: Sure. I would like my students to understand the relationship now between theme and character development and plot and setting and how all of those work to shape the piece.

Me: And why does any of that matter? Why should they learn that?

Educator: Well, like I said before, I want them to be able to pick up a play or story and make sense of what the author is trying to communicate so they can make some personal connections to it and hopefully make some more sense of their lives. Also, I hope they realize that literature is another way they can communicate with others. So, by learning the techniques of Shakespeare and others they can learn how to express themselves effectively and creatively. Maybe those should be my standards, making sense of the world and communicating effectively, or are those too broad?

Me: Those are too broad for standards. Those sound like your overall goals for your course. But you could not easily assess such goals in one or two measures. You want to break them down into several standards that capture the key components of your goals and that are amenable to assessment. So, let's go back to your statement about the relationship of theme to the other elements of literature. It's not that being able to identify a theme is a useless skill. But you want your students to go beyond that. How can we frame what you said as a standard?

Educator: How about

Students will explain the relationships between theme, character, setting ...

Do I need to list all the literary elements I cover?

Me: You could. Or, if that might change from one year to another you could say something like

Students will explain the relationships between several literary elements (e.g., theme, character, setting, plot)

Educator: You can do that in a standard?

Me: Yes, you can do anything you want in writing a standard as long as it captures significant learning you value and is written in a manner that can be assessed.

Educator: But there are some elements, like theme, that I would always want them to understand.

Me: Then you can say "several literary elements including theme, character, setting, and plot"

Educator: That's better. So, how about this?

Students will explain relationships between and among literary elements including character, plot, setting, theme, conflict and resolution and their influence on the effectiveness of the literary piece.

Me: Very nice! Is it realistic?

Educator: Yes, I think so.

Me: Is it something worth learning?

Educator: Definitely.

Me: Can you assess it?

Educator: Oh yes, there would be a lot of ways. So.... are we done?

Me: Yes. You have developed an excellent standard.

Educator: That was a lot of work.

Me: Yes. It is not easy to write good standards. But, after you have done a few the rest will come more easily.

Educator: (with a touch of sarcasm) Oh, sure.

Step 2: Select an Authentic Task

Note: Before you begin this section I would recommend you read the section on [Authentic Tasks](#) to learn about characteristics and types of authentic tasks.

► [Starting from Scratch: Look at Your Standards](#)

► [Starting from Scratch: Look at the Real World](#)

► **NEW** **Workshop: Creating an Authentic Task**

If you completed Step 1 (identify your standards) successfully, then the remaining three steps, particularly this one, will be much easier. With each step it is helpful to return to your goals and standards for direction. For example, imagine that one of your standards is

Students will describe the geographic, economic, social and political consequences of the Revolutionary War.

In Step 2, you want to find a way students can demonstrate that they are fully capable of meeting the standard. The language of a well-written standard can spell out what a task should ask students to do to demonstrate their mastery of it. For the above standard it is as simple as saying the task should ask students to *describe the geographic, economic, social and political consequences of the Revolutionary War*. That might take the form of an analytic paper you assign, a multimedia presentation students develop (individually or collaboratively), a debate they participate in or even an essay question on a test.

"Are those all authentic tasks?"

Yes, because each one a) asks students to construct their own responses and b) replicates meaningful tasks found in the real world.

"Even an essay question on a test? I thought the idea of Authentic Assessment was to get away from tests."

First, authentic assessment does not compete with traditional assessments like tests. Rather, they complement each other. Each typically serves different assessment needs, so a combination of the two is often appropriate. Second, if you read the section on **Authentic Tasks** I mentioned above (and I am beginning to doubt you did :-), then you will recall that essay questions fall near the border between traditional and authentic assessments. Specifically, essay questions are constructed-response items. That is, in response to a prompt, students construct an answer out of old and new knowledge. Since there is no one exact answer to these prompts, students are constructing new knowledge that likely differs slightly or significantly from that constructed by other students. Typically, constructed response prompts are narrowly conceived, delivered at or near the same time a response is expected and are limited in length. However, the fact that students must construct new knowledge means that at least some of their thinking must be revealed. As opposed to selected response items, the teachers gets to look inside the head a little with constructed response answers. Furthermore, explaining or analyzing as one might do in an essay answer replicates a real-world skill one frequently uses. On the other hand, answering a question such as

Which of the following is a geographical consequence of the Revolutionary War?

- a.
- b.
- c.
- d.

requires students to *select* a response, not *construct* one. And, circling a correct answer is not a significant challenge that workers or citizens commonly face in the real world.

So, yes, it can be that easy to construct an authentic assessment. In fact, you probably recognize that some of your current assessments are authentic or performance-based ones. Moreover, I am guessing that you feel you get a better sense of your students' ability to apply what they have learned through your authentic assessments than from your traditional assessments.

Starting from Scratch?: Look at your Standards

What if you do not currently have an authentic assessment for a particular standard? How do you create one from scratch? Again, start with your standard. What does it ask your students to do? A good authentic task would ask them to demonstrate what the standard expects of students. For example, the standard might state that students will

solve problems involving fractions using addition, subtraction, multiplication and division.

Teachers commonly ask students to do just that -- solve problems involving fractions. That is an authentic task.

See an example of the process of creating an authentic task from a standard in the [workshop](#) below.

Starting from Scratch?: Look at the Real World

But what if you want a more engaging task for your students? A second method of developing an authentic task from scratch is by asking yourself "where would they use these skills in the real world?" For computing with fractions teachers have asked students to follow recipes, order or prepare pizzas, measure and plan the painting or carpeting of a room, etc. Each of these tasks is not just an instructional activity; each can also be an authentic assessment.

See more [examples](#) of authentic tasks.

NEW Workshop: Creating an Authentic Task

In the "workshops" sprinkled throughout this website I will attempt to capture (and model) the process I follow when assisting someone or some group in developing standards or authentic tasks or rubrics. For this workshop, I will begin with a particular skill an imaginary educator would like to develop and assess in her second grade students, and we will work towards an authentic means of assessing the skill. You can "play along at home" by imagining how you would respond to the educator or to me.

Somewhere in Vienna (hey, it's my workshop; I'll host it where I like!)

Educator: I often get frustrated when my students constantly ask me whether they think their work is any good or not, or when they ask me if I think they are finished with some task. I want them to learn to judge those things for themselves. I need to teach more of that. But I have no idea how I would measure something like that. Is that really an authentic skill, and could I really assess it?

Me: No and no. Let's go have some [Sachertorte](#). Just kidding. First, is it authentic? Do you ever find yourself needing to reflect on your own work, to figure out what is working and what is not, to make changes when necessary, or to decide when you have finished something?

Educator: Of course. I do that all the time as a teacher, like when I am working on a lesson plan. I do that in a lot of situations, or I wouldn't get much better at whatever I am working on.

Me: That point is well supported by a recent article from Wiggins and McTighe (2006) entitled, "[Examining the teaching life](#)," in which they describe how educators can reflect upon their work "in light of sound principles about how learning works." So, it certainly is an authentic skill. Authentic tasks do not have to be large, complex projects. Most mental behaviors are small, brief "tasks" such as deciding between two choices, or interpreting a political cartoon, or finding a relationship between two or more concepts. Thus, many authentic tasks we give our students can and should be small and brief, whether they are for practicing some skill or assessing students on it.

Educator: But are second graders too young to evaluate their own work?

Me: No, teachers can and have begun developing this skill in kindergarteners. As with anything, start simple and small.

As you may know, considering how to assess such a skill in the classroom usually begins by referring to your standards. Did you write a standard addressing the skill you described?

Educator: Yes. In fact, I completed your absolutely fabulous "[Writing a good standard](#)" workshop. So, see what you think of what I came up with.

Students will evaluate their own work.

Is that okay? I know it is rather broad. I could have chosen more specific elements of self-assessment such as identifying errors in their work or judging if they have completed the assignment. But I want my students to begin acquiring all the skills of self-evaluation so I wrote the standard with that in mind.

Me: I think that is a reasonable standard. Your standard may be broad in some sense, but I notice that you are limiting it to evaluating the students' work, not their behavior. As you probably know, some teachers ask their students to evaluate their own *behavior* during the day. For example, students are asked to assess how well they are contributing to the class, staying on task, avoiding or resolving conflicts with others, etc. I think the scope of your standard is appropriate and manageable. So, let's go with that standard. If you need to change it as we consider the tasks you always can. Nothing in assessment is written in stone.

Now, second, can you assess it? "Evaluate" is an observable verb. But, what does "evaluate their own work" actually look like when people are doing it?

Educator: When I think of evaluating one's own work or self-assessing I think of things like

- judging the quality of one's work
- identifying one's strengths and weaknesses
- finding errors and correcting them when necessary

Me: Those are very good examples. Other ways of saying much the same thing include

- comparing one's work against specific criteria or standard (which is similar to judging its quality)
- or comparing it to past work or the work of others
- reflecting upon one's work:
 - does it meet the goal(s)?
 - in other words, have I finished yet?
 - where are there discrepancies between the goal(s) and one's current piece of work?
 - what do I need to improve?
 - am I making progress?

Notice in our list of skills that with the exception of *correcting them when necessary* all of the statements focus on identifying how well one is performing and *not* on the next step of identifying strategies for improvement or addressing one's weaknesses. Although correcting one's errors or devising strategies for improvement follows logically from identifying those errors or weaknesses, the two sets of skills can be considered, taught and assessed independently of each other. So, I think it makes sense for you to limit your focus to the first step of *evaluating one's work*. Given that, which of the evaluation skills do you want your students to develop?

Educator: All of them really.

Me: There is quite a bit of overlap or redundancy in the list we created. Can you consolidate those skills into two or three that you would like to focus on here?

Educator: Well, as I mentioned before, I would like my students to stop asking me or their parents or others all the time if their work is any good. Sometimes they will need to check with others. But, I want them to be able to determine if their work is any good for themselves, whether that means being able to compare their work against a set of criteria or a rubric I might give them or just knowing what "good" looks like for a particular task. Related to that, I would like my students to be able to judge when they are "done" with a task. Yes, I want them to recognize when the minimum requirements have been met, but I also want them to judge when they have produced something worthwhile.

Me: Very good. We should not have too much difficulty thinking of tasks you could assign your students that will indicate whether or not they are acquiring those skills.

Educator: First, I want to check on something: Just because I have a standard for something, do I have to assess it?

Me: Only the most essential understandings and skills should be captured in your standards. Thus, if it is important enough to include in your standards you will want to know if your students are meeting those goals. You will want to assess it. On the other hand, there may be skills that you would like to promote or encourage in your students, but you don't consider them critical. So, you don't have to assess them. However, if this is a skill you would really like to teach and develop in your students...

Educator: It is...

Me: Then you will want to assess it, which brings us back to your original question. How can you assess the skill described in your standard: Students will evaluate their own work? Let's start with the first skill you described: Judging the quality of their own work. To get you started, here are a few possible options:

- applying the rubric for a specific task to their own work on that task ([click here](#) to see some elementary level examples)
- applying a generic self-assessment rubric applicable to most tasks to their work on a specific task
- applying a generic self-assessment rubric applicable to most tasks to a collection of student work over a period of time
- identifying strengths and/or weaknesses in their work on a task or across a collection of work
- answering some open-ended questions about their work such as
 - what do you like about your work on _____?
 - what did you find difficult/easy?
 - what still needs improvement?
 - what do you need more help with?
 - what do you still need to learn more about for this task?
 - what did you discover about yourself as you worked on this task?
 - if you had 24 more hours to work on this task, what changes would you make?

So, pick one of these and flesh it out to give me a task that would work in your class. You have 30 seconds.

Educator: What?!? Okay, um, how about ... I got it! I borrowed the [Fairy Tale Letter](#) task from your Toolbox developed by Debra Crooks and Kate Trtan. They created a good rubric for the task. So, I could do the following with my students:

- 1) Assign my students the Fairy Tale Letter task with a certain time or date for completion of a draft. I will give them the rubric before they begin the task.
- 2) When the students have written a draft of the letter, I will ask them to review the rubric.
- 3) Then I will ask them to review their letter draft.
- 4) Next, I will ask them to circle the descriptor that best fits their letter for each criterion.
- 5) Then I will collect their drafts and rubrics on which they circled the descriptors.
- 6) I will judge how well they have applied the rubric to their drafts.
- 7) Finally, I will return their drafts and rubrics so they can complete the letter.

Me: That's a good start.

Educator: Uh oh. I know what that means when you say "a good start."

Me: I mean that you have described a very good framework for assessing self-assessment in this manner. I just think your task needs a little tweaking. In fact, we need to do the very thing you are asking your students to do: Evaluate the quality of your work. How do you think you judge the quality of a task you have created, adapted or borrowed?

Educator: The task should align with the standard. So, first, I want to make sure I am really assessing whether or not students are evaluating their own work. Of course, it looks good to me -- I wrote it! So, how can I try to more objectively evaluate the task?

Me: A good strategy for evaluating a task is to imagine possible student performance on the task and see if you can really determine whether the standard was met or not. For example, if this is the rubric,

Criteria	5	3	1
Parts of a letter	Correctly used all parts of a letter	Omitted one part of a letter	Omitted more than one part of a letter
Number of sentences	At least five sentences	Used four sentences	Used fewer than four sentences

Sentence structure	Complete sentences with correct mechanics	Sentences are incomplete or mechanics errors	Sentences are incomplete and mechanics errors
Voice	Used character voice throughout entire letter	Used character voice throughout most of the letter	Used character voice throughout little of the letter

imagine a student scored himself a 5 (correctly used all parts of a letter) for the "parts of a letter" criterion, a 5 for number of sentences, a 3 for sentence structure, and a 5 for voice on his draft. When you look at the student's draft, you score him a 3, 5, 3, 3. What have you learned about how well this student can evaluate his own work?

Educator: Well, I can tell that the student recognized that most of the parts of the letter were there, but he missed one part. Also, he correctly realized that he included five sentences. He appears to be aware that there were some incomplete sentences or mechanical errors, but I cannot tell which errors he identified. Finally, the student did not seem to realize that his character lost its voice in a few places.

Me: That's a possible interpretation of the student's ratings. Is it also possible that the student just guessed and happened to agree with you on some criteria by chance?

Educator: I guess that's possible, too. How could I tell if he just guessed?

Me: I was about to ask you that. If you were there with the student, what would you do to find out?

Educator: I would just ask him: Why did you circle "Correctly used all parts of a letter" for that criterion?

Me: Then you can make that question part of your assessment. But, before we consider how you might incorporate that formally into your assessment, let's go back to the way you originally described it. Simply asking your students to apply the rubric to their drafts is a good task in itself. It may not tell you formally whether or not they are meeting the standard, but it serves as good practice for this skill. And with any skill, you would want to give them feedback on it. So, you could give them your ratings on the rubric and ask them to compare them with their own. With second graders, it may not be very helpful just to see your ratings without some assistance. But you could

- meet with some or all of your students individually to share your ratings and ask some questions like you mentioned (they can also be invited to ask you questions about how you arrived at your ratings so you can model that thought process)
- assign the students to pairs in which they help each other compare their ratings to yours to see if they can figure out why there is a discrepancy for one or more of the criteria
- ask them to pick one criterion where your rating differed from theirs and then carefully review their draft for that criterion again

By simply asking your students to apply the rubric and examining their ratings you will get some sense of how well they are judging their own work. You may notice certain patterns such as they all seem to be able to determine if they have included enough sentences, but they are quite poor at judging whether their character has used a consistent voice. So, as an informal assessment, I think your task (and its many possible variations) should give you some useful information and provide some good practice in the skill of self-assessment.

However, if you want to draw more valid inferences about how well the students are meeting your standard, you will need to collect evidence that more clearly indicates how well your students are evaluating their own work. Earlier, you said you could help determine if students were just guessing when they applied the rubric by asking them follow-up questions. How might we include such questions as part of the task?

Educator: For a more formal assessment, I could give the students the rubric at the top of a sheet with a few questions at the bottom. After they apply the rubric to their drafts, the students could be directed to answer the questions. For example, I could give them one of the following sheets:

1)

- Review the rubric below.
- Then review your draft of the fairy tale letter.
- Circle the descriptors (such as "correctly used all parts of a letter") that best describe your draft.
- Finally, answer the questions below the rubric.

The Rubric

For each criterion in the rubric above, explain why you circled the level (5, 3, or 1) you did.

Parts of a letter

Number of sentences

Sentence structure

Voice

2)

- Review the rubric below.
- Then review your draft of the fairy tale letter.
- Circle the descriptors (such as "correctly used all parts of a letter") that best describe your draft.
- Finally, answer the questions below the rubric.

The Rubric

Which rating that you just circled in the rubric do you feel **most** confident about? Tell me why.

Which rating that you just circled in the rubric do you feel **least** confident about? Tell me why.

3)

- Review the rubric below.
- Then review your draft of the fairy tale letter.
- Circle the descriptors (such as "correctly used all parts of a letter") that best describe your draft.
- Finally, answer the question below the rubric.

The Rubric

Pick one criterion that you rated the lowest in the rubric above. What could you do in your letter draft to move you up to the next level in the rubric for that criterion?

Me: Those are very good questions. I would feel more confident about assessing a student's ability to evaluate his work if, in addition to completing the rubric, he also had to answer one or more of those questions. You will have *made his thinking visible* so you can more easily discern whether he arrived at his answer through guessing or through genuine reflection on his level of performance.

Engaging in such self-assessment, particularly with some thoughtful reflection, is not an easy task by any means, and particularly not for second graders. Skill development requires careful scaffolding. So, we must assume that administering an assessment such as one of these for your students would come only after considerable practice with the skill. Furthermore, practice should follow significant teacher modeling. For example, you could write a fairy tale letter, intentionally including some stronger and weaker parts. Then, you would walk through the rubric with your students to illustrate how to apply the rubric. You could model it yourself, or you could invite their participation in the process. Similarly, asking students to apply a rubric to someone else's work, whether another student's in the class or a mock sample you provide them, should also provide good practice.

Alternatively, some teachers provide students with samples of what specific descriptors might look like. For example, you might share examples of what a 5 or a 3 or a 1 looks like for the criterion of Voice for the Fairy Tale Letter task.

Of course, even if students take the task seriously and attempt to fairly judge their work, they still may have great difficulty doing so. For example, one of the criteria in the above rubric is "sentence structure," and applying that criterion means judging if the sentences are complete and the mechanics are free from errors. That is not always easy for good writers; how will weak writers know? An interesting article by [Dunning et al. \(2003\)](#) entitled, "[Why people fail to recognize their own incompetence,](#)" describes research finding that "...poor performers are doubly cursed: Their lack of skill deprives them not only of the ability to produce correct responses, but also of the expertise necessary to surmise that they are not producing them" (p. 83). Thus, before many of our students can effectively evaluate their own work we need to equip them with the meta-cognitive skills of thinking about how they would accomplish that task.

In other words, *how* would good or weak writers determine if their writing contains mechanical errors? If they cannot do that, they cannot yet apply that criterion in the rubric. So, another question we might ask a second grader or a sixth grader or a high school senior when applying a rubric to a task is

- How will you determine which level of that criterion applies to your work?

Of course, the easy answer to that is "ask my teacher," and we are back where we started this whole discussion! But, if we teach and model the meta-cognitive strategies underlying good self-assessment, then eventually we should get some intelligent answers to that question, and better self-assessment, and better performance.

So, what do you think? Could you feasibly assess how well your students could evaluate their own work?

Educator: I think so. At least I am much more confident about it than when we started. It will definitely take a lot of practice and feedback and reflection.

Me: Is it worth the time?

Educator: Definitely. Instead of spending all that time asking my students to learn to apply criteria to their work and then giving them assessments on it, I could have them devote more time to working on their fairy tale letter, for example. But, in the long run, I believe they will produce better work if they can confidently critique it themselves, they will acquire a truly valuable skill that they can apply to almost every facet of their lives, and I may even save time if they become more efficient at producing good work.

Me: You sold me. But, you know what? We are not done yet. Eons ago, or whenever we started this conversation, you also said you would like your students to acquire a related ability: The ability to judge when their work is "done." We will try to keep this brief, but let's see if we can come up with a task or two to assess that skill.

Educator: Okay. Your turn. You've got 15 seconds. Go!

Me: What?!?!? My...mind...is...blank...oh, here we go. Before students turn in a particular assignment, and, perhaps, after reviewing the assignment rubric, give them one of the following sheets:

Have you completed the requirements of Assignment X?
Yes No
If not, what do you still need to do to complete the assignment?

Have you completed Assignment X <u>well</u> ?
Yes No
If you said Yes, how do you know it is finished and it is done well?
If you said No, how do you know it is not finished or not yet done well?
or
If you said No, what still needs to be improved?
or
If you said No, how will you know when it is done well?



Educator: At this point, my students would have a hard time answering those questions. Yet, as you said, you have to start somewhere. I could definitely model answers to those questions, and I would give my students plenty of practice, feedback and opportunity for reflection on these skills. The tasks we created should help me teach my students self-assessment skills and provide me a tool for assessing the standard. So, are we done here?

Me: One more thing... We created some possible tasks, but for a formal assessment of the skill you would need some way to score your students' performance.

Educator: A rubric?

Me: That's one possibility. Authentic assessments are not required to include a rubric; some do, some don't. But we will save rubrics for a rubric workshop. To get your thinking started in that direction I just want to ask you to briefly identify a few of the criteria you would look for in your students' efforts on these tasks. What would be the characteristics of good performance on your first self-assessment task that you might measure?

Educator: I would probably look for the following indicators:

- Did the students select the appropriate descriptors in the rubric for their drafts?
- For the first two sheets, did they provide reasonable justification for their choices?
- Or, for the last sheet, did their answer indicate a good grasp of their deficiencies?

Me: Very nice. We're done! Oh, could you grab that sheet of paper on the table.

Educator: What is it?

Me: It's my Sachertorte rubric. I like to hit four or five Viennese restaurants or hotels and compare. It's research!

Step 3: Identify the Criteria for the Task

▶ [Examples of Criteria](#)

▶ [Characteristics of a Good Criterion](#)

▶ [How Many Criteria do you Need for a Task?](#)

▶ [Time for a Quiz!](#)

Criteria: *Indicators of good performance on a task*

In Step 1, you identified what you want your students to know and be able to do. In Step 2, you selected a task (or tasks) students would perform or produce to demonstrate that they have met the standard from Step 1. For Step 3, you want to ask "What does good performance on this task look like?" or "How will I know they have done a good job on this task?" In

answering those questions you will be identifying the *criteria* for good performance on that task. You will use those criteria to evaluate how well students completed the task and, thus, how well they have met the standard or standards.

Examples

Example 1: Here is a *standard* from the [Special Education](#) collection of [examples](#):

The student will conduct banking transactions.

The *authentic task* this teacher assigned to students to assess the standard was to

make deposits, withdrawals or cash checks at a bank.

To identify the *criteria* for good performance on this task, the teacher asked herself "what would good performance on this task look like?" She came up with seven essential characteristics for successful completion of the task:

- Selects needed form (deposit, withdrawal)
- Fills in form with necessary information
- Endorses check
- Locates open teller
- States type of transaction
- Counts money to be deposited to teller
- Puts money received in wallet

If students meet these criteria then they have performed well on the task and, thus, have met the standard or, at least, provided some evidence of meeting the standard.

Example 2: This comes from the [Mathematics](#) collection. There were *six standards* addressed to some degree by this authentic task. The standards are: Students will be able to

- measure quantities using appropriate units, instruments, and methods;
- setup and solve proportions;
- develop scale models;
- estimate amounts and determine levels of accuracy needed;
- organize materials;
- explain their thought process.

The *authentic task* used to assess these standards in a geometry class was the following:

Rearrange the Room

You want to rearrange the furniture in some room in your house, but your parents do not think it would be a good idea. To help persuade your parents to rearrange the furniture you are going to make a two dimensional scale model of what the room would ultimately look like.

Procedure:

- 1) You first need to measure the dimensions of the floor space in the room you want to rearrange, including the location and dimensions of all doors and windows. You also need to measure the amount of floor space occupied by each item of furniture in the room. These dimensions should all be explicitly listed.
- 2) Then use the given proportion to find the scale dimensions of the room and all the items.
- 3) Next you will make a scale blueprint of the room labeling where all windows and doors are on poster paper.
- 4) You will also make scale drawings of each piece of furniture on a cardboard sheet of paper, and these models need to be cut out.
- 5) Then you will arrange the model furniture where you want it on your blueprint, and tape them down.
- 6) You will finally write a brief explanation of why you believe the furniture should be arranged the way it is in your model.

Your models and explanations will be posted in the room and the class will vote on which setup is the best.

Finally, the *criteria* which the teacher identified as indicators of good performance on the Rearrange the Room task were:

- accuracy of calculations;
- accuracy of measurements on the scale model;
- labels on the scale model;
- organization of calculations;
- neatness of drawings;
- clear explanations.

(But *how well* does a student have to perform on each of these criteria to do well on the task? We will address that question in [Step 4: Create the Rubric.](#))

You may have noticed in the second example that some of the standards and some of the criteria sounded quite similar. For example, one standard said students will be able to *develop scale models*, and two of the criteria were *accuracy of measurements on the scale model* and *labels on the scale model*. Is this redundant? No, it means that your criteria are aligned with your standards. You are actually measuring on the task what you said you valued in your standards.

Characteristics of a Good Criterion

So, what does a good criterion (singular of criteria) look like? It should be

- a clearly stated;
- brief;
- observable;
- statement of behavior;
- written in language students understand.

Additionally, make sure each criterion is distinct. Although the criteria for a single task will understandably be related to one another, there should not be too much overlap between them. Are you really looking for different aspects of performance on the task with the different criteria, or does one criterion simply rephrase another one? For example, the following criteria might be describing the same behavior depending on what you are looking for:

- interpret the data
- draw a conclusion from the data

Another overlap occurs when one criterion is actually a subset of another criterion. For example, the first criterion below probably subsumes the second:

- presenter keeps the audience's attention
- presenter makes eye contact with the audience

Like standards, criteria should be shared with students *before* they begin a task so they know the teacher's expectations and have a clearer sense of what good performance should look like. Some teachers go further and involve the students in identifying appropriate criteria for a task. The teacher might ask the students "What characteristics does a good paper have?" or "What should I see in a good scale model?" or "How will I (or anyone) know you have done a good job on this task?"

How Many Criteria do you Need for a Task?

Of course, I am not going to give you an easy answer to that question because there is not one. But, I can recommend some guidelines.

- **Limit the number of criteria; keep it to the essential elements of the task.** This is a guideline, not a rule. On a major, complex task you might choose to have 50 different attributes you are looking for in a good performance. That's fine. But, generally, assessment will be more feasible and meaningful if you focus on the important characteristics of the task. Typically, you will have fewer than 10 criteria for a task, and many times it might be as few as three or four.
- **You do not have to assess everything on every task.** For example, you might value correct grammar and spelling in all writing assignments, but you do not have to look for those criteria in every assignment. You have made it clear to your students that you expect good grammar and spelling in every piece of writing, but you only check for it in some of them. That way, you are assessing those characteristics in the students' writing and you are sending the message that you value those elements, but you do not take the time of grading them on every assignment.
- **Smaller, less significant tasks typically require fewer criteria.** For short homework or in-class assignments you might only need a quick check on the students' work. Two or three criteria might be sufficient to judge the understanding or application you were after in that task. Less significant tasks require less precision in your assessment than larger, more comprehensive tasks that are designed to assess significant progress toward multiple standards.

Ask. Ask yourself; you have to apply the criteria. Do they make sense to you? Can you distinguish one from another? Can you envision examples of each? Are they all worth assessing?

Ask your students. Do they make sense to them? Do they understand their relationship to the task? Do they know how they would use the criteria to begin their work? To check their work?

Ask your colleagues. Ask those who give similar assignments. Ask others who are unfamiliar with the subject matter to get a different perspective if you like.

If you have assigned a certain task before, review previous student work. Do these criteria capture the elements of what you considered good work? Are you missing anything essential?

Time for a Quiz!

Do you think you could write a good criterion now? Do you think you would know a good one when you saw one? Let's give you a couple small tasks:

Task 1: Write three criteria for a good employee at a fast-food restaurant. (There would likely be more than three, but as a simple check I do not need to ask for more than three. Assessments should be meaningful and manageable!)

Task 2: I have written three criteria for a good employee below. I intentionally wrote two clear criteria (I hope) and one vague one. Can you find the vague one among the three? Are the other two good criteria? (Yes, I wrote them so of course I think they are good criteria. But I will let you challenge my authority just this once :-)

- the employee is courteous
- the employee arrives on time
- the employee follows the sanitary guidelines

What do you think? In my opinion, the first criterion is vague and the latter two are good criteria. Of course, evaluating criteria is a subjective process, particularly for those you wrote yourself. So, before I explain my rationale I would reiterate the advice above of checking your criteria with others to get another opinion.

To me, the statement "the employee is courteous" is too vague. Courteous could mean a lot of different things and could mean very different things to different people. I would think the employer would want to define the behavior more specifically and with more clearly observable language. For example, an employer might prefer:

- the employee greets customers in a friendly manner

That is a more observable statement, but is that all there is to being courteous? It depends on what you want. If that is what the employer means by courteous then that is sufficient. Or, the employer might prefer:

- the employee greets customers in a friendly manner *and* promptly and pleasantly responds to their requests

"Is that one or two criteria?" It depends on how detailed you want to be. If the employer wants a more detailed set of criteria he/she can spell out each behavior as a separate criterion. Or, he/she might want to keep "courteous" as a single characteristic to look for but define it as two behaviors in the criterion. There is a great deal of flexibility in the number and specificity of criteria. There are few hard and fast rules in any aspect of assessment development. You need to make sure the assessment fits your needs. An employer who wants a quick and dirty check on behavior will create a much different set of criteria than one who wants a detailed record.

The second criterion above, the employee arrives on time, is sufficiently clear. It cannot obviously name a specific time for arriving because that will change. But if the employer has identified the specific time that an employee should arrive than "arrive on time" is very clear. Similarly, if the employer has made clear the sanitary guidelines, then it should be clear to the employees what it means to "follow the guidelines."

"Could I include some of that additional detail in my criteria or would it be too wordy?" That is up to you. However, criteria are more communicable and manageable if they are brief. The employer could include some of the definition of courteous in the criterion statement such as

- the employee is courteous (i.e., the employee greets customers in a friendly manner *and* promptly and pleasantly responds to their requests)

However, it is easier to state the criterion as "the employee is courteous" while explaining to the employees exactly what behaviors that entails. Whenever the employer wants to talk about this criterion with his/her employees he can do it more simply with this brief statement. We will also see how rubrics are more manageable (coming up in Step 4) if the criteria are brief.

"Can I have sub-criteria in which I break a criterion into several parts and assess each part separately?" Yes, although that might be a matter of semantics. Each "sub-criterion" could be called a separate criterion. But I will talk about how to handle that in the next section "Step 4: Create the Rubric."

Step 4: Create the Rubric

► [Creating an Analytic Rubric](#)

► [Creating a Holistic Rubric](#)

► [Final Step: Checking Your Rubric](#)

► [Workshop: Writing a Good Rubric](#)

Note: Before you begin this section I would recommend that you read the section on [Rubrics](#) to learn about the characteristics of a good rubric.

In Step 1 of creating an authentic assessment, you identified what you wanted your students to know and be able to do -- your standards.

In Step 2, you asked how students could demonstrate that they had met your standards. As a result, you developed authentic tasks they could perform.

In Step 3, you identified the characteristics of good performance on the authentic task -- the criteria.

Now, in Step 4, you will finish creating the authentic assessment by constructing a rubric to measure student performance on the task. To build the rubric, you will begin with the set of criteria you identified in Step 3. As mentioned before, keep the number of criteria manageable. You do not have to look for everything on every assessment.

Once you have identified the criteria you want to look for as indicators of good performance, you next decide whether to consider the criteria analytically or holistically. (See [Rubrics](#) for a description of these two types of rubrics.)

Creating an Analytic Rubric

In an *analytic rubric* performance is judged separately for each criterion. Teachers assess how well students meet a criterion on a task, distinguishing between work that effectively meets the criterion and work that does not meet it. The next step in creating a rubric, then, is deciding how fine such a distinction should be made for each criterion. For example, if you are judging the amount of eye contact a presenter made with his/her audience that judgment could be as simple as did or did not make eye contact (two levels of performance), never, sometimes or always made eye contact (three levels), or never, rarely, sometimes, usually, or always made eye contact (five levels).

Generally, it is better to start small with fewer levels because it is usually harder to make more fine distinctions. For eye contact, I might begin with three levels such as never, sometimes and usually. Then if, in applying the rubric, I found that some students seemed to fall in between never and sometimes, and never or sometimes did not adequately describe the students' performance, I could add a fourth (e.g., rarely) and, possibly, a fifth level to the rubric.

In other words, there is some trial and error that must go on to arrive at the most appropriate number of levels for a criterion. (See the Rubric Workshop below to see more detailed decision-making involved in selecting levels of performance for a sample rubric.)

Do I need to have the same number of levels of performance for each criterion within a rubric?

No. You could have five levels of performance for three criteria in a rubric, three levels for two other criteria, and four levels for another criterion, all within the same rubric. Rubrics are very flexible tools. There is no need to force an unnatural judgment of performance just to maintain standardization within the rubric. If one criterion is a simple either/or judgment and another criterion requires finer distinctions, then the rubric can reflect that variation.

Here are some examples of rubrics with varying levels of performance.....

Do I need to add descriptors to each level of performance?

No. *Descriptors* are recommended but not required in a rubric. As described in *Rubrics*, descriptors are the characteristics of behavior associated with specific levels of performance for specific criteria. For example, in the following portion of an elementary science rubric, the criteria are 1) observations are thorough, 2) predictions are reasonable, and 3) conclusions are based on observations. Labels (limited, acceptable, proficient) for the different levels of performance are also included. Under each label, for each criterion, a descriptor (in brown) is included to further explain what performance *at that level* looks like.

Criteria	Limited	Acceptable	Proficient
made good observations	observations are absent or vague	most observations are clear and detailed	all observations are clear and detailed
made good predictions	predictions are absent or irrelevant	most predictions are reasonable	all predictions are reasonable
appropriate conclusion	conclusion is absent or inconsistent with observations	conclusion is consistent with most observations	conclusion is consistent with observations

As you can imagine, students will be more certain what is expected to reach each level of performance on the rubric if descriptors are provided. Furthermore, the more detail a teacher provides about what good performance looks like on a task the better a student can approach the task. Teachers benefit as well when descriptors are included. A teacher is likely to be more objective and consistent when applying a descriptor such as "*most observations are clear and detailed*" than when applying a simple label such as "*acceptable*." Similarly, if more than one teacher is using the same rubric, the

specificity of the descriptors increases the chances that multiple teachers will apply the rubric in a similar manner. When a rubric is applied more consistently and objectively it will lead to greater reliability and validity in the results.

Assigning point values to performance on each criterion

As mentioned above, rubrics are very flexible tools. Just as the number of levels of performance can vary from criterion to criterion in an analytic rubric, points or value can be assigned to the rubric in a myriad of ways. For example, a teacher who creates a rubric might decide that certain criteria are more important to the overall performance on the task than other criteria. So, one or more criteria can be weighted more heavily when scoring the performance. For example, in a rubric for solo auditions, a teacher might consider five criteria: (how well students demonstrate) vocal tone, vocal technique, rhythm, diction and musicality. For this teacher, musicality might be the most important quality that she has stressed and is looking for in the audition. She might consider vocal technique to be less important than musicality but more important than the other criteria. So, she might give musicality and vocal technique more weight in her rubric. She can assign weights in different ways. Here is one common format:

Rubric 1: Solo Audition

	0	1	2	3	4	5	weight
vocal tone							
vocal technique							x2
rhythm							
diction							
musicality							x3

In this case, placement in the 4-point level for vocal tone would earn the student four points for that criterion. But placement in the 4-point box for vocal technique would earn the student 8 points, and placement in the 4-point box for musicality would earn the student 12 points. The same weighting could also be displayed as follows:

Rubric 2: Solo Audition

	NA	Poor	Fair	Good	Very Good	Excellent
vocal tone	0	1	2	3	4	5
vocal technique	0	2	4	6	8	10
rhythm	0	1	2	3	4	5
diction	0	1	2	3	4	5
musicality	0	3	6	9	12	15

In both examples, musicality is worth three times as many points as vocal tone, rhythm and diction, and vocal technique is worth twice as much as each of those criteria. Pick a format that works for you and/or your students. There is no "correct" format in the layout of rubrics. So, choose one or design one that meets your needs.

Yes, but do I need equal intervals between the point values in a rubric?

No. Say it with me one more time -- rubrics are flexible tools. Shape them to fit your needs, not the other way around. In other words, points should be distributed across the levels of a rubric to best capture the value you assign to each level of performance. For example, points might be awarded on an oral presentation as follows:

Rubric 3: Oral Presentation

Criteria	never	sometimes	always
----------	-------	-----------	--------

makes eye contact	0	3	4
volume is appropriate	0	2	4
enthusiasm is evident	0	2	4
summary is accurate	0	4	8

In other words, you might decide that at this point in the year you would be pleased if a presenter makes eye contact "sometimes," so you award that level of performance most of the points available. However, "sometimes" would not be as acceptable for level of volume or enthusiasm.

Here are some more examples of rubrics illustrating the flexibility of number of levels and value you assign each level.

Rubric 4: Oral Presentation

Criteria	never	sometimes	usually
makes eye contact	0	2	4
volume is appropriate	0		4
enthusiasm is evident	0		4
summary is accurate	0	4	8

In the above rubric, you have decided to measure volume and enthusiasm at two levels -- never or usually -- whereas, you are considering eye contact and accuracy of summary across three levels. That is acceptable if that fits the type of judgments you want to make. Even though there are only two levels for volume and three levels for eye contact, you are awarding the same number of points for a judgment of "usually" for both criteria. However, you could vary that as well:

Rubric 5: Oral Presentation

Criteria	never	sometimes	usually
makes eye contact	0	2	4
volume is appropriate	0		2
enthusiasm is evident	0		2
summary is accurate	0	4	8

In this case, you have decided to give less weight to volume and enthusiasm as well as to judge those criteria across fewer levels.

So, do not feel bound by any format constraints when constructing a rubric. The rubric should best capture what you value in performance on the authentic task. The more accurately your rubric captures what you want your students to know and be able to do the more valid the scores will be.

Creating a Holistic Rubric

In a *holistic rubric*, a judgment of how well someone has performed on a task considers all the criteria together, or holistically, instead of separately as in an analytic rubric. Thus, each level of performance in a holistic rubric reflects behavior across all the criteria. For example, here is a holistic version of the oral presentation rubric above.

Rubric 6: Oral Presentation (Holistic)

Oral Presentation Rubric

Mastery

- usually makes eye contact
- volume is always appropriate
- enthusiasm present throughout presentation
- summary is completely accurate

Proficiency

- usually makes eye contact
- volume is usually appropriate
- enthusiasm is present in most of presentation
- only one or two errors in summary

Developing

- sometimes makes eye contact
- volume is sometimes appropriate
- occasional enthusiasm in presentation
- some errors in summary

Inadequate

- never or rarely makes eye contact
- volume is inappropriate
- rarely shows enthusiasm in presentation
- many errors in summary

An obvious, potential problem with applying the above rubric is that performance often does not fall neatly into categories such as mastery or proficiency. A student might always make eye contact, use appropriate volume regularly, occasionally show enthusiasm and include many errors in the summary. Where you put that student in the holistic rubric? Thus, it is recommended that the use of holistic rubrics be limited to situations when the teacher wants to:

- make a quick, holistic judgment that carries little weight in evaluation, or
- evaluate performance in which the criteria cannot be easily separated.

Quick, holistic judgments are often made for homework problems or journal assignments. To allow the judgment to be quick and to reduce the problem illustrated in the above rubric of fitting the best category to the performance, the number of criteria should be limited. For example, here is a possible holistic rubric for grading homework problems.

Rubric 7: Homework Problems

Homework Problem Rubric

++ (3 pts.)

- most or all answers correct, AND

- most or all work shown

+ (1 pt.)

- at least some answers correct, AND
- at least some but not most work shown

- (0 pts.)

- few answers correct, OR
- little or no work shown

Although this homework problem rubric only has two criteria and three levels of performance, it is not easy to write such a holistic rubric to accurately capture what an evaluator values *and* to cover all the possible combinations of student performance. For example, what if a student got all the answers correct on a problem assignment but did not show any work? The rubric covers that: the student would receive a (-) because "little or no work was shown." What if a student showed all the work but only got some of the answers correct? That student would receive a (+) according to the rubric. All such combinations are covered. But does giving a (+) for such work reflect what the teacher values? The above rubric is designed to give equal weight to correct answers and work shown. If that is not the teacher's intent then the rubric needs to be changed to fit the goals of the teacher.

All of this complexity with just two criteria -- imagine if a third criterion were added to the rubric. So, with holistic rubrics, limit the number of criteria considered, or consider using an analytic rubric.

Final Step: Checking Your Rubric

As a final check on your rubric, you can do any or all of the following before applying it.

- Let a colleague review it.
- Let your students review it -- is it clear to them?
- Check if it aligns or matches up with your standards.
- Check if it is manageable.
- Consider imaginary student performance on the rubric.

By the last suggestion I mean to imagine that a student had met specific levels of performance on each criterion (for an analytic rubric). Then ask yourself if that performance translates into the score that you think is appropriate. For example, on Rubric 3 above, imagine a student scores

- "sometimes" for eye contact (3 pts.)
- "always" for volume (4 pts.)
- "always" for enthusiasm (4 pts.)
- "sometimes" for summary is accurate (4 pts.)

That student would receive a score of 15 points out of a possible 20 points. Does 75% (15 out of 20) capture that performance for you? Perhaps you think a student should not receive that high of a score with only "sometimes" for the summary. You can adjust for that by increasing the weight you assign that criterion. Or, imagine a student apparently put a lot of work into the homework problems but got few of them correct. Do you think that student should receive some credit? Then you would need to adjust the holistic homework problem rubric above. In other words, it can be very helpful to play out a variety of performance combinations before you actually administer the rubric. It helps you see the forest through the trees.

Of course, you will never know if you really have a good rubric until you apply it. So, do not work to perfect the rubric before you administer it. Get it in good shape and then try it. Find out what needs to be modified and make the appropriate changes.

Okay, does that make sense? Are you ready to create a rubric of your own? Well, then come into my workshop and we will build one together. I just need you to wear these safety goggles. Regulations. Thanks.

(For those who might be "tabularly challenged" (i.e., you have trouble making tables in your word processor) or would just like someone else to make the rubric into a tabular format for you, there are websites where you enter the criteria and levels of performance and the site will produce the rubric for you. One such site is the [Rubric Builder](#).)

Examples

For example standards: <http://jonathan.mueller.faculty.noctrl.edu/toolbox/examples/standardsexamples.htm>

For example tasks & rubrics: <http://jonathan.mueller.faculty.noctrl.edu/toolbox/examples/authentictaskexamples.htm>

For example portfolios: <http://jonathan.mueller.faculty.noctrl.edu/toolbox/examples/portfoliosexamples.htm>