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 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**

http://jfmueller.faculty.noctrl.edu/toolbox/howdoyoudoit.htm
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**
- **Step 2: Select an Authentic Task**
- **Step 3: Identify the Criteria for the Task**
- **Step 4: Create the Rubric**
Before I can effectively teach or assess students, I need to determine what they should know and be able to do. In other words, I need a good set of standards. Or do I need goals? Or objectives? Standards, goals and objectives are all descriptions of what students should know and be able to do. So, how are they different? I distinguish standards from other statements of student performance primarily along two dimensions: 1) breadth of coverage and 2) feasibility of measurement and observation.

**Breadth of Coverage**

Starting at the top of the above diagram, the mission statement of schools or districts or states is typically the broadest statement of what students are intended to know and be able to do when they graduate. In roughly 50 words or less, mission statements attempt to communicate to all constituencies the purposes of education in that institution. An example of a mission statement might be:

"All students at Mueller School will become effective communicators, collaborators and problem-solvers."

Unfortunately, mission statements just make good wall-hangings in many schools or districts. That is a missed opportunity. A clearly written, purposeful statement can serve as an excellent starting point for curriculum development, instruction and assessment. Furthermore, a good mission statement can provide a useful guide against which progress can be compared to determine if it is following a consistent, productive path.

For example, if Mueller School adopted the above statement, it would design all curriculum in a manner to promote effective communication, collaboration and problem-solving. Disciplinary content would not be forgotten. Rather, a clear focus would develop around teaching students to communicate about mathematics, collaborate in the construction of new knowledge and solve problems specific to science, social studies, the fine arts, etc. As a teacher at that school I would always ask myself if the lesson I had planned or the curricular framework we developed would promote such knowledge and skills. Thus, a good mission statement would serve as a focal point to initiate development as well as a check for progress.

As stated above, mission statements are very brief, broad statements. To flesh them out further schools often identify a set of goals which more specifically, yet still broadly, define expectations for students. The goals also communicate the school’s or district’s focus for its educational plan.

Goals are typically subdivided further to identify standards. Whereas goals are often written broadly enough to cross grade levels and content areas, standards, particularly those that are
content-based, tend to be specific to one or a few grade levels and one content area, and may be written at the level of a unit in curricular planning. However, many state and national K-12 standards are written with the graduating senior in mind. To provide guidance for prior grades, benchmark standards are written which describe what progress third or fifth or eighth graders should have made toward a particular standard.

Moving down the pyramid above, the statements of what students should know and be able to do become more narrow and, consequently, more numerous within a curriculum. The most specific and numerous is the objective. Objectives are typically written at the level of the lesson plan, with one or more objectives for each lesson.

**Feasibility of Measurement**

The four types of statements presented in the pyramid can also be differentiated by whether or not they are amenable to assessment. Goals and mission statements are typically written to share a broad vision, not to serve as benchmarks to be measured. Thus, their language does not usually make them amenable to assessment. On the other hand, standards and objectives are written with measurement in mind. Consequently, those statements need to describe student behavior that is observable.

So, why is there a section of this authentic assessment web site devoted to standards and not one on mission statements, goals or objectives? Although the term standard has been around the field of education for a long time, it has become more prominent in recent years as the authentic assessment movement has taken off. I believe it has become more prominent than the other statements of behavior in the movement for two reasons. First, like objectives, standards are amenable to assessment, a necessary requirement to guide task design. Second, the broader nature of standard versus objective is consistent with authentic assessment’s emphasis on complex, integrative authentic tasks that typically span more than one class period, more than one topic and sometimes even more than one discipline.

Thus, good authentic assessment development begins with identifying a set of standards for your students. State and national efforts at standards-writing have typically focused on the content of the disciplines. But what about critical thinking skills, problem solving abilities, collaborative skills and personal development? These highly valued skills are not easily incorporated into content standards and, thus, are often omitted or given insufficient attention. Yet, the standards should capture what we most value and most want our students to learn. So, we should consider including these other skills in our standards. To do so, it may be helpful to distinguish content standards from other types. To see how, look at

►Types of Standards

---

Copyright 2016, Jon Mueller. Professor of Psychology, North Central College, Naperville, IL. Comments, questions or suggestions about this website should be sent to the author, Jon Mueller, at jfmueller@noctrl.edu.
**Authentic Tasks**

**Authentic Task:** An assignment given to students designed to assess their ability to apply standard-driven knowledge and skills to real-world challenges.

In other words, a task we ask students to perform is considered authentic when 1) students are asked to construct their own responses rather than select from ones presented and 2) the task replicates challenges faced in the real world. (Of course, other definitions abound.)

If I were teaching you how to play golf, I would not determine whether you had met my standards by giving you a multiple-choice test. I would put you out on the golf course to "construct your own responses" in the face of real-world challenges. Similarly, in school we are ultimately less interested in how much information students can acquire than how well they can use it. Thus, our most meaningful assessments ask students to perform authentic tasks.

However, these tasks are not just assessments. 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.

**Characteristics of Authentic Tasks**

Another way that authentic assessment is commonly distinguished from traditional assessment is in terms of their defining attributes. Of course, traditional assessments as well as authentic assessments vary considerably in the forms they take. But, typically, along the continuums of attributes listed below, traditional assessments fall more towards the left end of each continuum and authentic assessments fall more towards the right end.

<table>
<thead>
<tr>
<th>Traditional</th>
<th>Authentic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Selecting a Response</td>
<td>Performing a Task</td>
</tr>
<tr>
<td>Contrived</td>
<td>Real-life</td>
</tr>
<tr>
<td>Recall/Recognition</td>
<td>Construction/Application</td>
</tr>
<tr>
<td>Teacher-structured</td>
<td>Student-structured</td>
</tr>
<tr>
<td>Indirect Evidence</td>
<td>Direct Evidence</td>
</tr>
</tbody>
</table>

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.

**Types of Authentic Tasks**

I have used the term *traditional assessment* on this site to refer to the many tests that are commonly administered to assess the acquisition of knowledge and skills. Tests usually consist of selected-response items (see below) and, occasionally, some constructed-response items. In contrast, *authentic assessments* include tasks such as performances, products and constructed-response items that typically require more direct application of knowledge and skills. These types of tasks are described below along with common examples of each.

**Selected-response**

In response to a prompt, students select an answer from among those given or from memory or from allowable study aids. Typically, no new knowledge is constructed; students simply recall or recognize information required to select the appropriate response. Examples include

- Multiple-choice tests
- True-false
- Matching
- Fill-in-the-blank
- Label a diagram

**Constructed Response**

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. Examples include

(product-like):

- Short-answer essay questions
- "Show your work"
- Ordering decimals
Limericks and rubric
Concept maps; another example / rubric
Writing a topic sentence
Identifying a theme
Making predictions
Brief summaries; another example
Peer editing
Figural representation (e.g., Venn diagram; web / rubric)
Journal response; literary journal reflections
Homework reflections; article reflections / rubric
Evaluating work of others; another example; another example
Self-assessment; another example / rubric
Self and group evaluation
Goal setting; another example / reflection
Question generation; another example
Explain your solution

(performancelike):
Typing test
Complete a step of science lab
Measure objects
Conducting bank transactions
Utilizing library services
Computer catalog search
On demand, construct a short musical, dance or
dramatic response
On demand, exhibit an athletic skill
Reading fluently
Conferences
Participation (and self-assessment)

Product
In response to a prompt (assignment) or series of prompts, students construct a substantial,
tangible product that reveals their understanding of certain concepts and skills and/or their ability
to apply, analyze, synthesize or evaluate those concepts and skills. It is similar to a constructed-
response item in that students are required to construct new knowledge and not just select a
response. However, product assessments typically are more substantial in depth and length, more
broadly conceived, and allow more time between the presentation of the prompt and the student
response than constructed-response items. Examples include

Essays, stories, or poems
Ballads
Obituaries
Satirical pieces
Metaphors
School rules
Research reports; another example
Annotated bibliographies
Works cited pages
Reading strategies and rubric
Projects / rubric; another example / rubric; another example
Literary analysis; another example; another example
Character analysis; another example
Diction analysis
Advertisement analysis
Biography/Autobiography analysis
Argument analysis / rubric
Analyzing primary sources
Analysis of painting
Film analysis
Geometric analysis
Article reviews
Book reviews / rubric
Case study / rubric
Speech critiques
Extended journal responses
Identification of goals
Reading guides
Feudal contracts / rubric
Art exhibit or portfolio
Models; another example
Constructing objects
Floor plans
Musical compositions
Photo compositions
Design an advertisement
Design an experiment
Lab reports; another example
Surveys
Data recordings
Graphing of data
Data analysis; another example; another example
Analysis of statistical use in media / rubric
Real-world problem solutions; another example / rubric
Logical sequences
Error analysis
Planning for a task
Preparing for a discussion
Proposals and criteria
Road trip directions
Map construction / rubric
Road trip budget
Scavenger hunt
Newspapers
Newscasts; another example
Editorials; another example
Peer editing / rubric
Posters; another example; another example / rubric
Collages
Pamphlets; another example
Brochures; another example / rubric
Magazine covers
Bulletin boards
Videos / rubric
Podcasts
Games; another example; another example
Comic strips
Books; Booklets
Timelines; another example / rubric
Issue awareness campaigns
Letter writing; persuasive letter writing; complaint letter
Advice letter; letter to Congress; letter to Emperor

Performance

In response to a prompt (assignment) or series of prompts, students construct a performance that reveals their understanding of certain concepts and skills and/or their ability to apply, analyze, synthesize or evaluate those concepts and skills. It is similar to a constructed-response item in that students are required to construct new knowledge and not just select a response. However, performances typically are more substantial in depth and length, more broadly conceived, and allow more time between the presentation of the prompt and the student response than constructed-response items. Examples include

Conducting an experiment
Musical auditions; group auditions
Conducting an ensemble / rubric
Conduct band rehearsal / rubric
Create musical arrangement / rubric
Dance or dramatic performances
Dramatic readings
Skits
Role-plays / handout
Talk show performances; another example
Book talks
Debates; another example / rubric
Panel discussions
Fishbowl discussions
Coffee shop conversation
Athletic competitions
Oral presentations; another example; another example
Teaching/explaining
Speeches
Interviews
Self-introduction
Cooperative group behavior; another example

Also, see
Examples of Authentic Tasks

How to Select an Authentic Task
Rubrics

Descriptors

Why Include Levels of Performance?

Analytic Versus Holistic Rubrics

How Many Levels of Performance Should I Include in my Rubric?

Rubric: A scoring scale used to assess student performance along a task-specific set of criteria

Authentic assessments typically are criterion-referenced measures. That is, a student’s aptitude on a task is determined by matching the student’s performance against a set of criteria to determine the degree to which the student’s performance meets the criteria for the task. To measure student performance against a pre-determined set of criteria, a rubric, or scoring scale, is typically created which contains the essential criteria for the task and appropriate levels of performance for each criterion. For example, the following rubric (scoring scale) covers the research portion of a project:

<table>
<thead>
<tr>
<th>Criteria</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Sources</td>
<td>1-4</td>
<td>5-9</td>
<td>10-12</td>
</tr>
<tr>
<td>Historical Accuracy</td>
<td>Lots of historical inaccuracies</td>
<td>Few inaccuracies</td>
<td>No apparent inaccuracies</td>
</tr>
<tr>
<td>Organization</td>
<td>Can not tell from which source information came</td>
<td>Can tell with difficulty where information came from</td>
<td>Can easily tell which sources info was drawn from</td>
</tr>
<tr>
<td>Bibliography</td>
<td>Bibliography contains very little information</td>
<td>Bibliography contains most relevant information</td>
<td>All relevant information is included</td>
</tr>
</tbody>
</table>

As in the above example, a rubric is comprised of two components: criteria and levels of performance. Each rubric has at least two criteria and at least two levels of performance. The criteria, characteristics of good performance on a task, are listed in the left-hand column in the rubric above (number of sources, historical accuracy, organization and bibliography). Actually, as is common in rubrics, the author has used shorthand for each criterion to make it fit easily into the table. The full criteria are statements of performance such as "include a sufficient number of sources" and "project contains few historical inaccuracies."

For each criterion, the evaluator applying the rubric can determine to what degree the student has met the criterion, i.e., the level of performance. In the above rubric, there are three levels of performance for each criterion. For example, the project can contain lots of historical inaccuracies, few inaccuracies or no inaccuracies.

Finally, the rubric above contains a mechanism for assigning a score to each project. (Assessments and their accompanying rubrics can be used for purposes other than evaluation and, thus, do not have to have points or grades attached to them.) In the second-to-left column a weight is assigned each criterion. Students can receive 1, 2 or 3 points for "number of sources." But historical accuracy, more important in this teacher’s mind, is weighted three times (x3) as heavily. So, students can receive 3, 6 or 9 points (i.e., 1, 2 or 3 times 3) for the level of accuracy in their projects.

Descriptors

The above rubric includes another common, but not a necessary, component of rubrics -- descriptors. Descriptors spell out what is expected of students at each level of performance for each criterion. In the above example, "lots of historical inaccuracies," "can tell with difficulty where information came from" and "all relevant information is included" are descriptors. A descriptor tells students more precisely what performance looks like at each level and how their work may be
distinguished from the work of others for each criterion. Similarly, the descriptors help the teacher more precisely and consistently distinguish between student work.

Many rubrics do not contain descriptors, just the criteria and labels for the different levels of performance. For example, imagine we strip the rubric above of its descriptors and put in labels for each level instead. Here is how it would look:

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Poor (1)</th>
<th>Good (2)</th>
<th>Excellent (3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Sources</td>
<td>x1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Historical Accuracy</td>
<td>x3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Organization</td>
<td>x1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bibliography</td>
<td>x1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

It is not easy to write good descriptors for each level and each criterion. So, when you first construct and use a rubric you might not include descriptors. That is okay. You might just include the criteria and some type of labels for the levels of performance as in the table above. Once you have used the rubric and identified student work that fits into each level it will become easier to articulate what you mean by "good" or "excellent." Thus, you might add or expand upon descriptors the next time you use the rubric.

### Why Include Levels of Performance?

#### Clearer expectations

As mentioned in Step 3, it is very useful for the students and the teacher if the criteria are identified and communicated prior to completion of the task. Students know what is expected of them and teachers know what to look for in student performance. Similarly, students better understand what good (or bad) performance on a task looks like if levels of performance are identified, particularly if descriptors for each level are included.

#### More consistent and objective assessment

In addition to better communicating teacher expectations, levels of performance permit the teacher to more consistently and objectively distinguish between good and bad performance, or between superior, mediocre and poor performance, when evaluating student work.

#### Better feedback

Furthermore, identifying specific levels of student performance allows the teacher to provide more detailed feedback to students. The teacher and the students can more clearly recognize areas that need improvement.

### Analytic Versus Holistic Rubrics

For a particular task you assign students, do you want to be able to assess how well the students perform on each criterion, or do you want to get a more global picture of the students' performance on the entire task? The answer to that question is likely to determine the type of rubric you choose to create or use: Analytic or holistic.

#### Analytic rubric

Most rubrics, like the Research rubric above, are analytic rubrics. An analytic rubric articulates levels of performance for each criterion so the teacher can assess student performance on each criterion. Using the Research rubric, a teacher could assess whether a student has done a poor, good or excellent job of "organization" and distinguish that from how well the student did on "historical accuracy."

#### Holistic rubric

In contrast, a holistic rubric does not list separate levels of performance for each criterion. Instead, a holistic rubric assigns a level of performance by assessing performance across multiple criteria as a whole. For example, the analytic research rubric above can be turned into a holistic rubric:
In the analytic version of this rubric, 1, 2 or 3 points is awarded for the number of sources the student included. In contrast, number of sources is considered along with historical accuracy and the other criteria in the use of a holistic rubric to arrive at a more global (or holistic) impression of the student work. Another example of a holistic rubric is the "Holistic Critical Thinking Scoring Rubric" (in PDF) developed by Facione & Facione.

**When to choose an analytic rubric**

Analytic rubrics are more common because teachers typically want to assess each criterion separately, particularly for assignments that involve a larger number of criteria. It becomes more and more difficult to assign a level of performance in a holistic rubric as the number of criteria increases. For example, what level would you assign a student on the holistic research rubric above if the student included 12 sources, had lots of inaccuracies, and whose bibliography contained most relevant information? As student performance increasingly varies across criteria it becomes more difficult to assign an appropriate holistic category to the performance. Additionally, an analytic rubric better handles weighting of criteria. How would you treat "historical accuracy" as more important a criterion in the holistic rubric? It is not easy. But the analytic rubric handles it well by using a simple multiplier for each criterion.

**When to choose a holistic rubric**

So, when might you use a holistic rubric? Holistic rubrics tend to be used when a quick or gross judgment needs to be made. If the assessment is a minor one, such as a brief homework assignment, it may be sufficient to apply a holistic judgment (e.g., check, check-plus, or no-check) to quickly review student work. But holistic rubrics can also be employed for more substantial assignments. On some tasks it is not easy to evaluate performance on one criterion independently of performance on a different criterion. For example, many writing rubrics (see example) are holistic because it is not always easy to disentangle clarity from organization or content from presentation. So, some educators believe a holistic or global assessment of student performance better captures student ability on certain tasks. (Alternatively, if two criteria are nearly inseparable, the combination of the two can be treated as a single criterion in an analytic rubric.)

**How Many Levels of Performance Should I Include in my Rubric?**

There is no specific number of levels a rubric should or should not possess. It will vary depending on the task and your needs. A rubric can have as few as two levels of performance (e.g., a checklist) or as many as ... well, as many as you decide is appropriate. (Some do not consider a checklist a rubric because it only has two levels -- a criterion was met or it wasn't. But because a checklist does contain criteria and at least two levels of performance, I include it under the category of rubrics.) Also, it is not true that there must be an even number or an odd number of levels. Again, that will depend on the situation.

To further consider how many levels of performance should be included in a rubric, I will separately address analytic and holistic rubrics.

**Analytic rubrics**

Generally, it is better to start with a smaller number of levels of performance for a criterion and then expand if necessary. Making distinctions in student performance across two or three broad categories is difficult enough. As the number of levels increases, and those judgments become finer and finer, the likelihood of error increases.

Thus, start small. For example, in an oral presentation rubric, amount of eye contact might be an important criterion. Performance on that criterion could be judged along three levels of performance: never, sometimes, always.

<table>
<thead>
<tr>
<th>makes eye contact with audience</th>
<th>never</th>
<th>sometimes</th>
<th>always</th>
</tr>
</thead>
</table>

Although these three levels may not capture all the variation in student performance on the criterion, it may be sufficient discrimination for your purposes. Or, at the least, it is a place to start. Upon applying the three levels of performance, you might discover that you can effectively group your students' performance in these three categories. Furthermore, you might discover that the labels of never, sometimes and always sufficiently communicates to your students the degree to which they can improve on making eye contact.
On the other hand, after applying the rubric you might discover that you cannot effectively discriminate among student performance with just three levels of performance. Perhaps, in your view, many students fall in between never and sometimes, or between sometimes and always, and neither label accurately captures their performance. So, at this point, you may decide to expand the number of levels of performance to include never, rarely, sometimes, usually and always.

| makes eye contact | never | rarely | sometimes | usually | always |

There is no “right” answer as to how many levels of performance there should be for a criterion in an analytic rubric; that will depend on the nature of the task assigned, the criteria being evaluated, the students involved and your purposes and preferences. For example, another teacher might decide to leave off the “always” level in the above rubric because “usually” is as much as normally can be expected or even wanted in some instances. Thus, the “makes eye contact” portion of the rubric for that teacher might be:

| makes eye contact | never | rarely | sometimes | usually |

So, I recommend that you begin with a small number of levels of performance for each criterion, apply the rubric one or more times, and then re-examine the number of levels that best serve your needs. I believe starting small and expanding if necessary is preferable to starting with a larger number of levels and shrinking the number because rubrics with fewer levels of performance are normally

- easier and quicker to administer
- easier to explain to students (and others)
- easier to expand than larger rubrics are to shrink

The fact that rubrics can be modified and can reasonably vary from teacher to teacher again illustrates that rubrics are flexible tools to be shaped to your purposes. To read more about the decisions involved in developing a rubric, see the chapter entitled, "Step 4: Create the Rubric."

**Holistic rubrics**

Much of the advice offered above for analytic rubrics applies to holistic rubrics as well. Start with a small number of categories, particularly since holistic rubrics often are used for quick judgments on smaller tasks such as homework assignments. For example, you might limit your broad judgments to:

- satisfactory
- unsatisfactory
- not attempted

or

- check-plus
- check
- no check

or even just

- satisfactory (check)
- unsatisfactory (no check)

Of course, to aid students in understanding what you mean by “satisfactory” or “unsatisfactory” you would want to include descriptors explaining what satisfactory performance on the task looks like.

Even with more elaborate holistic rubrics for more complex tasks I recommend that you begin with a small number of levels of performance. Once you have applied the rubric you can better judge if you need to expand the levels to more effectively capture and communicate variation in student performance.

To read more about the decisions involved in developing rubrics, see

acija

[Step 4: Create a Rubric]