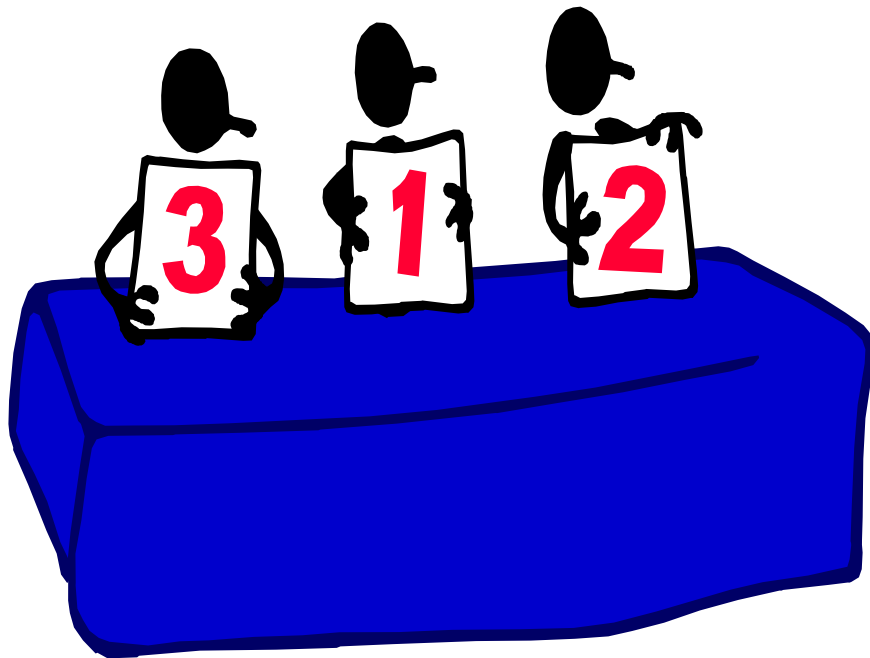


Evaluating and Anchoring Student Performance



presented by

Jay McTighe

Educational Author and Consultant

E-mail: jay@mctighe-associates.com

Website: jaymctighe.com

Four Categories of Criteria

Content – refers to the appropriateness and relative sophistication of the understanding, knowledge and skill employed.

Quality – refers to the overall quality, craftsmanship and rigor of the work.

Process – refers to the quality and appropriateness of the procedures, methods, and approaches used, prior to and during performance.

Result – refers to the impact, success or effectiveness of performance, given the purpose(s) and audience.

Example – Cooking a Meal

Here is an example in which all four types of criteria might be used to evaluate a meal in nine different ways:

Content

1. meal reflects knowledge of food, cooking, situation, and diners' needs and tastes
2. meal contains the appropriate, fresh ingredients
3. meal reflects sophisticated flavors and pairings

Quality

4. meal is presented in aesthetically appealing manner
5. all dishes are cooked to taste

Process

6. meal is efficiently prepared, using appropriate techniques
7. the two cooks collaborated effectively

Result

8. meal is nutritious
9. meal is pleasing to all guests

Note: While these four categories reflect common types of criteria, I do not mean to suggest that you must use all four types for each and every performance task. Rather, you should select the criterion types that are appropriate for the goals being assessed through the task and for which you want to provide feedback to learners.

Four Categories of Criteria

Content – refers to the appropriateness and relative sophistication of the understanding, knowledge and skill employed.

- *Was the work accurate?*
- *Did the product reveal deep understanding?*
- *Were the answers appropriately supported?*
- *Was the work thorough?*
- *Were the arguments of the essay cogent?*
- *Was the hypothesis plausible and on target?*
- *In sum: Was the content appropriate to the task, accurate, and supported?*

Quality – refers to the overall quality, craftsmanship and rigor of the work.

- *Was the speech organized?*
- *Was the paper mechanically sound?*
- *Was the chart clear and easy to follow?*
- *Did the story build and flow smoothly?*
- *Was the dance graceful?*
- *Were the graphics original?*
- *In sum: Was the performance or product of high quality?*

Process – refers to the quality and appropriateness of the procedures, methods, and approaches used, prior to and during performance.

- *Was the performer methodical?*
- *Was proper procedure followed?*
- *Was the planning efficient and effective?*
- *Did the reader/problem solver employ apt strategies?*
- *Did the group work collaboratively and effectively?*
- *In sum: Was the approach sound?*

Result – refers to the impact, success or effectiveness of performance, given the purpose(s) and audience.

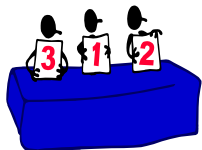
- *Was the desired result achieved?*
- *Was the problem solved?*
- *Was the client satisfied?*
- *Was the audience engaged and informed?*
- *Was the dispute resolved?*
- *Did the speech persuade?*
- *Did the paper open minds to new possibilities?*
- *In sum: Was the work effective?*

Categories of Performance Criteria

By what criteria should understanding performances be assessed? The challenge in answering is to ensure that we assess what is *central* to the understanding, not just what is easy to score. In addition, we need to make sure that we identify the *separate* traits of performance (e.g. a paper can be well-organized but not informative and vice versa) to ensure that the student gets specific and valid feedback. Finally, we need to make sure that we consider the different *types* of criteria (e.g. the quality of the *understanding* vs. the quality of the *performance* in which it is revealed).

Four types of performance criteria (with sample indicators)

content	process	quality	result
Describes the degree of knowledge of factual information or understanding of concepts, principles, and processes.	Describes the degree of skill/proficiency. Also refers to the effectiveness of the process or method used.	Describes the degree of quality evident in products and performances.	Describes the overall impact and the extent to which goals, purposes, or results are achieved.
accurate appropriate authentic complete correct credible explained justified important in-depth insightful logical makes connections precise relevant sophisticated supported thorough valid	careful clever coherent collaborative concise coordinated effective efficient flawless followed process logical/reasoned mechanically correct methodical meticulous organized planned purposeful rehearsed sequential skilled	attractive competent creative detailed extensive focussed graceful masterful organized polished proficient precise neat novel rigorous skilled stylish smooth unique well-crafted	beneficial conclusive convincing decisive effective engaging entertaining informative inspiring meets standards memorable moving persuasive proven responsive satisfactory satisfying significant useful understood



Criterion-Based Evaluation Tools

There are various types of scoring tools that teachers can use to evaluate student performance. The selection of the appropriate tool is determined by the answer to the following questions:

Key Questions

- What is the **purpose** of this performance task or assignment (*diagnostic, formative, summative*)?
- What **evaluation tool** is most appropriate given the assessment purpose?
 - performance list holistic rubric analytic rubric
 - generic task specific
- What is the **range of the scale**?
- **Who will use** the evaluation tool (*teachers, external scorers, students, others*)?
What format will be most useful for the chosen purpose, user and audience?

Note: If students are involved, the tool should be written in understandable ‘kid language.’

TYPES OF CRITERION-BASED EVALUATION TOOLS

	SCORING RUBRIC	PERFORMANCE LIST
	Holistic : Analytic	Analytic
Generic		
Task-Specific		

PERFORMANCE LIST for Graphic Display of Data (elementary level)

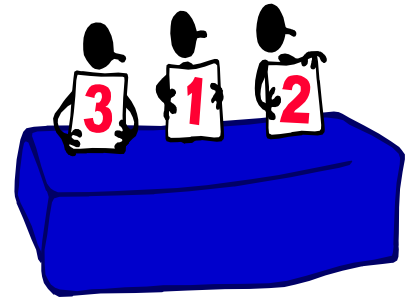
Key Criteria	Points Possible	Self	Other	Teacher
1. The graph contains a title that tells what the data shows.	<input type="text"/>	_____	_____	_____
2. All parts of the graph (units of measurement, rows, etc.) are correctly labelled.	<input type="text"/>	_____	_____	_____
3. All data is accurately represented on the graph.	<input type="text"/>	_____	_____	_____
4. The graph is neat and easy to read.	<input type="text"/>	_____	_____	_____
Total	<input type="text"/>	_____	_____	_____

Performance lists offer a practical means of judging student performance based upon identified criteria. A performance list consists of a set of criterion elements or traits and a rating scale. The rating scale is quite flexible, ranging from 3 to 100 points.

Teachers can assign points to the various elements, in order to “weight” certain elements over others (e.g., accuracy counts more than neatness) based on the relative importance given the achievement target. The lists may be configured to easily convert to conventional grades. For example, a teachers could assign point values and weights that add up to 25, 50 or 100 points, enabling a straightforward conversion to a district or school grading scale (e.g., 90-100 = A, 80-89 = B, and so on). When the lists are shared with students in advance, they provide a clear performance target, signaling to students what elements should be present in their work.

Despite these benefits, performance lists do not provided detailed descriptions of *performance levels*. Thus, despite identified criteria, different teachers using the same performance list may rate the same student’s work quite differently.

SCORING RUBRICS



Definition

Rubrics are criterion-based evaluation tools used in conjunction with “open-ended” performance tasks and projects, which do not have a single, “correct” answer or solution process.

Two general types of rubrics – holistic and analytic – are widely used to judge student products and performances. A holistic rubric provides an overall impression of a student’s work. Holistic rubrics yield a *single* score or rating for a product or performance. An analytic rubric divides a product or performance into distinct traits or dimensions and judges each separately. Since an analytic rubric rates each of the identified traits independently, a separate score is provided for each.

A third type of rubric -- longitudinal -- describes growth along a fixed, novice-expert continuum, in which each level represents a key benchmark on the road to exit-level performance. These longitudinal rubrics provide a basis for designing backward from mastery performance so that teachers and learners at all levels know where they stand along a developmental continuum against exit-level performance goals. Longitudinal rubrics are not tied to any particular performance or assessment task. Rather, they enable teachers, parents, and learners to chart progress toward desired accomplishments.

Purpose

Effective rubrics:

- clearly define criteria for judging student performance;
- promote more consistent evaluation of student performance;
- help clarify instructional goals and serve as teaching targets;
- provide specific feedback to learners and teachers;
- help students focus on the important dimensions of a product or performance;
- support criterion-based assessment

Note: The criteria within a rubric should be directly linked to the targeted standards or learning outcomes, not simply focus on the surface features of student products or performances.

Constructing a Performance List

(example - oral presentation)

KEY QUESTIONS






















- What are the **key traits**, elements, or dimensions that will be evaluated?
- How many **score points** (scale) will be needed? (Checklists only need a binary scale – yes or no – when used to evaluate the presence or absence of elements.)

☞ Teachers should review and discuss the identified elements and the scale with students prior to using the performance list for self/peer/teacher evaluation.

Performance List for	<i>oral presentation</i>		
	Possible Points	Points Earned	
Key Traits:		<i>self</i>	<i>teacher</i>
• <u><i>topic explained and supported</i></u>	<u>30</u>	<u> </u>	<u> </u>
• <u><i>well organized</i></u>	<u>25</u>	<u> </u>	<u> </u>
• <u><i>effective visual display</i></u>	<u>25</u>	<u> </u>	<u> </u>
• <u><i>effective volume</i></u>	<u>5</u>	<u> </u>	<u> </u>
• <u><i>effective rate of speech</i></u>	<u>5</u>	<u> </u>	<u> </u>
• <u><i>appropriate inflection</i></u>	<u>5</u>	<u> </u>	<u> </u>
• <u><i>effective posture</i></u>	<u>5</u>	<u> </u>	<u> </u>
Totals	100	<input style="width: 50px; height: 30px;" type="text"/>	<input style="width: 50px; height: 30px;" type="text"/>

Performance List for Narrative Writing

Primary Level

	Terrific	O.K.	Needs Work
1. I have an interesting setting and characters for my story.			
2. The problem in my story will be clear to my readers.			
3. My story events are in order.			
4. The solution will be clear to my readers.			
5. I used many describing words to tell what is happening.			
6. My words “paint a picture.”			
7. I have a title that goes with my story.			

What will you try to do better the next time you write a story?

Holistic Rubric for Graphic Display of Data

3	All data is accurately represented on the graph. All parts of the graph (units of measurement, rows, etc.) are correctly labelled. The graph contains a title that clearly tells what the data shows. The graph is very neat and easy to read.
2	All data is accurately represented on the graph OR the graph contains minor errors. All parts of the graph are correctly labelled OR the graph contains minor inaccuracies. The graph contains a title that suggests what the data shows. The graph is generally neat and readable.
1	The data is inaccurately represented, contains major errors, OR is missing. Only some parts of the graph are correctly labelled OR labels are missing. The the title does not reflect what the data shows OR the title is missing. The graph is sloppy and difficult to read.

A holistic rubric provides an overall impression of a student's work. Holistic rubrics yield a *single* score or rating for a product or performance. Holistic rubrics are well suited to judging simple products or performances, such as a student's response to an open-ended test prompt. They provide a quick snapshot of overall quality or achievement, and are thus often used in large-scale assessment contexts (national, state or district levels) to evaluate a large number of student responses. Holistic rubrics are also effective for judging the "impact" of a product or performance (e.g., to what extent was the essay persuasive? did the play entertain?).

Despite these advantages, holistic rubrics have limitations. They do not provide a detailed analysis of the strengths and weaknesses of a product or performance. Since a single score is generally inadequate for conveying to students what they have done well and what they need to work on to improve, they are less effective at providing specific feedback to students.

A second problem with holistic rubrics relates to the interpretation and use of their scores. For instance, two students can receive the same score for vastly different reasons. Does an overall rating of "3" on a 4-point holistic writing rubric mean that a student has demonstrated strong idea development ("4") and weak use of conventions ("2"), or vice-versa? Without more specific feedback than a score or rating, it is difficult for the student to know exactly what to do to improve.

Holistic Rubric for Reading – Comprehension of Key Ideas and Details *(grades 4-5)*

Score Point 3

The student response an accurate analysis of what the text says explicitly and inferentially and references the text explicitly to support the analysis, showing full comprehension of complex ideas expressed in the text(s).

Score Point 2

The student response provides a mostly accurate analysis of what the text says explicitly and inferentially and references the text to support the analysis, showing comprehension of ideas expressed in the text(s).

Score Point 1

The student response provides a minimally accurate analysis of what the text says and may reference the text showing limited comprehension of ideas expressed in the text(s).

Score Point 0

The student response provides an inaccurate analysis or no analysis of the text, showing little to no comprehension of ideas expressed in the text(s).

Source: PARCC – Partnership for Assessment of Readiness for College and Careers

Analytic Rubric for Graphic Display of Data

	title	labels	accuracy	neatness
weights –				
3	The graph contains a title that clearly tells what the data shows. <input type="checkbox"/>	All parts of the graph (units of measurement, rows, etc.) are correctly labelled. <input type="checkbox"/>	All data is accurately represented on the graph. <input type="checkbox"/>	The graph is very neat and easy to read. <input type="checkbox"/>
2	The graph contains a title that suggests what the data shows. <input type="checkbox"/>	Some parts of the graph are inaccurately labelled. <input type="checkbox"/>	Data representation contains minor errors. <input type="checkbox"/>	The graph is generally neat and readable. <input type="checkbox"/>
1	The the title does not reflect what the data shows OR the title is missing. <input type="checkbox"/>	Only some parts of the graph are correctly labelled OR labels are missing. <input type="checkbox"/>	The data is inaccurately represented, contains major errors, OR is missing. <input type="checkbox"/>	The graph is sloppy and difficult to read. <input type="checkbox"/>

An analytic rubric divides a product or performance into distinct traits or dimensions and judges each separately. Since an analytic rubric rates each of the identified traits independently, a separate score is provided for each.

Analytic rubrics are better suited to judging complex performances (e.g., research process) involving several significant dimensions. As evaluation tools, they provide more specific information or feedback to students, parents and teachers about the strengths and weaknesses of a performance. Teachers can use the information provided by analytic evaluation to target instruction to particular areas of need. From an instructional perspective, analytic rubrics help students come to better understand the nature of quality work since they identify the important dimensions of a product or performance.

However, analytic rubrics are typically more time-consuming to learn and apply. Since there are several traits to be considered, analytic scoring may yield lower inter-rater reliability (degree of agreement among different judges) than holistic scoring. Thus, analytic scoring may be less desirable for use in large-scale assessment contexts, where speed and reliability are necessary.

Generic Analytic Speaking Rubric for World Languages

	Comprehensibility	Fluency	Pronunciation	Vocabulary	Language Control
4	Responses readily comprehensible, requiring no interpretation on the part of the listener.	Speech continuous with few pauses or stumbling.	Accurate pronunciation enhances communication.	Rich use of vocabulary enhances communication.	Accurate control of basic language structures.
3	Responses comprehensible, requiring minimal interpretation on the part of the listener.	Some hesitation but manages to continue and complete thoughts.	Infrequent mispronunciations do not interfere with communication.	Adequate and accurate use of vocabulary for this level enhances communication.	Generally accurate control of basic language structures.
2	Responses mostly comprehensible, requiring interpretation on the part of the listener.	Speech choppy and/or slow with frequent pauses; few or no incomplete thoughts.	Mispronunciations sometimes interfere with communication.	Inadequate and/or inaccurate use of vocabulary sometimes interferes w/ communication.	Emerging use of basic language structures.
1	Responses barely comprehensible.	Speech halting and uneven with long pauses or incomplete thoughts.	Frequent mispronunciations greatly interfere with communication.	Inadequate and/or inaccurate use of vocabulary greatly interferes with communication.	Inadequate and/or inaccurate use of basic language structures.

Source: Fairfax County, VA Public Schools <http://www.fcps.edu/DIS/OHSICS/forlang/PALS/rubrics/>

Generic Rubric for Collaboration

Works towards the achievement of group goals.

- 4 Actively helps identify group goals and works hard to meet them.
- 3 Communicates commitment to the group goals and effectively carries out assigned roles.
- 2 Communicates a commitment to the group goals but does not carry out assigned roles.
- 1 Does not work toward group goals or actively works against them.

Demonstrates effective interpersonal skills.

- 4 Actively promotes effective group interaction and the expression of ideas and opinions in a way that is sensitive to the feelings and knowledge base of others.
- 3 Participates in group interaction without prompting. Expresses ideas and opinions in a way that is sensitive to the feelings and knowledge base of others.
- 2 Participates in group interaction with prompting or expresses ideas and opinions without considering the feelings and knowledge base of others.
- 1 Does not participate in group interaction, even with prompting, or expresses ideas and opinions in a way that is insensitive to the feelings or knowledge base of others.

Contributes to group maintenance.

- 4 Actively helps the group identify changes or modifications necessary in the group process and works toward carrying out those changes.
- 3 Helps identify changes or modifications necessary in the group process and works toward carrying out those changes.
- 2 When prompted, helps identify changes or modifications necessary in the group process, or is only minimally involved in carrying out those changes.
- 1 Does not attempt to identify changes or modifications necessary in the group process, even when prompted, or refuses to work toward carrying out those changes.

Effectively performs a variety of roles within a group.

- 4 Effectively performs multiple roles within the group.
- 3 Effectively performs two roles within the group.
- 2 Makes an attempt to perform more than one role within the group but has little success with secondary roles.
- 1 Rejects opportunities or requests to perform more than one role in the group.

Source: Marzano, B., Pickering, D. and McTighe, J. (1993) *Assessing Outcomes: Performance Assessment based on the Dimensions of Learning Model*. Alexandria, VA: ASCD.

Task-Specific Rubric for a Science Investigation

Item 1 - Plan investigation (total possible points: 2)

- a) describes how the investigation will be conducted
- b) states what variables will be measured or observed; includes both solution time and temperature
- c) design provides control for other variables, or renders other variables irrelevant

Item 2 - Conduct investigation and record measurements in table

Response is scored for both the quality of the presentation and the quality of the data collection.

Quality of presentation (total possible points: 2)

- a) presents at least 2 sets of measurements in table.
- b) measurements are paired: dissolution time and temperature.
- c) labels table appropriately: data entries in columns identified by headings and/or units; units incorporated into headings or placed beside each measurement.

Quality of data (total possible points: 3)

- a) records solution time for at least three temperature points
- b) measurements are plausible: time and temperature (109 to 100 degrees)
- c) records solution times that decline as temperature increases

Item 3 - Draw conclusions about effect of temperature (total possible points: 2)

- a) conclusion is consistent with data table or other presentation of data
- b) describes relationship presented in the data

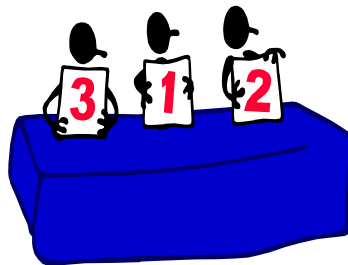
Item 4 - Explain conclusions (total possible points: 2)

- a) relates higher temperature to greater energy or speed of particles (atoms, molecules, etc.).
- b) makes connection between greater speed or energy of water molecules and the effect on the tablet (may be implicit).

Reviewing Your Rubric

In summary, the best rubrics:

1. evaluate student performances in terms of characteristics central to targeted standards and learning outcomes, not just the surface features of the task itself. Be careful not to over-emphasize the surface features of a particular product or performance (e.g., “colorful”, or “neat”) at the expense of the most important traits related to understanding (e.g., “thorough” or explanation with support”).
2. reflect the central features of performance, not just those which are easiest to see, count or score (e.g., “at least 4 footnotes” or “no misspellings”) at the expense of the most important traits (e.g., “accurate” or “effective”).
3. split independent criteria into separate traits. In other words, do not combine distinct traits, such as “very clear” and “very organized” in the same criterion, since an essay might be clear but not organized, and vice versa.
4. emphasize the result of the performance. Ultimately, authentic performance is about results – Was the paper persuasive?, ...the problem solved?, ...the story engaging?, ...the speech informative?, etc. In other words, the chosen criteria should always highlight the purpose of a task, in other words, as indicated by results-focused criteria. Be careful not to assess for mere compliance or process (i.e., “followed all the steps,” “worked hard”).
5. balance specific feedback on the task with reference back to general goals. Ultimately, a broad understanding matters more than performance on a unique and very specific task. However, the indicators need to be specific enough to provide useful feedback as well as reliable scoring of the particular task.



Rubric Refinement Process – Categorizing Student Work

The following six-step process for identifying performance criteria and using them as a basis for designing a scoring rubric. The procedure begins with sorting student work and then proceeds by looking at sample performance criteria from other places.

Step 1: Gather samples of student performance that illustrate the desired skill or understanding.

Choose as large and diverse a set of samples as possible.

Step 2: Sort student work into different stacks and write down the reasons.

For example, place the samples of student work into three piles: strong, middle and weak. As the student work is sorted, write down reasons for placing pieces in the various stacks. If a piece is placed in the “sophisticated” pile, describe its distinguishing features. What cues you that the work is sophisticated? What are you saying to yourself as you place a piece of work into a pile? What might you say to a student as you return this work? The qualities (attributes) that you identify reveal criteria. Keep sorting work until you are not adding anything new to your list of attributes.

Step 3: Cluster the reasons into traits or important dimensions of performance.

The sorting process used thus far in this exercise is “holistic.” Participants in this process end up with a list of comments for high, medium and low performance; any single student product gets only one overall score. Usually, during the listing of comments someone will say something to the effect that, “I had trouble placing this paper into one stack or another because it was strong on one trait but weak on another.” This brings up the need for analytical trait scoring systems; i.e., evaluating each student’s product or performance on more than one dimension.

Step 4: Write a definition of each trait.

These definitions should be “value neutral” – they describe what the trait is about, not what good performance looks like. (Descriptions of good performance on the trait are left to the “high” rating.)

Rubric Refinement Process

(continued)

Step 5: Find samples of student performance that illustrate each score point on each trait.

Find samples of student work which are good examples of strong, weak and mid range performance on each trait. These can be used to illustrate to students what to do and what “good” looks like. It’s important to have more than a single example. If you show students only a single example of what a good performance looks like, they are likely to imitate or copy it.

Step 6: Continuously Refine

Criteria and rubrics evolve with use. Try them out. You’ll probably find some parts of the rubric that work fine and some that don’t. Add and modify descriptions so that they communicate more precisely. Choose better sample papers that illustrate what you mean. Revise traits if you need to. When appropriate, let students help—this is a tool for learning.

Questions to consider when using a rubric to evaluate student work samples:	Possible rubric refinements:
<ul style="list-style-type: none"> • Have any important elements “fallen through the cracks”? Are important qualities that are evident in the best student work samples not specified in the rubric? 	<p><i>If so...</i> Add the missing element(s). Make sure that it (they) appear(s) consistently throughout the scale.</p>
<ul style="list-style-type: none"> • Is it difficult for reviewers to distinguish between two score points in the rubric? Are the distinctions between score points unclear or indistinguishable? 	<p><i>If so...</i> Consider shrinking the scale (e.g., from 6 to 5 points) so that the distinctions between levels are significant and readily determined.</p>
<ul style="list-style-type: none"> • Are raters asking to use + or – symbols next to the score points for some samples? 	<p><i>If so...</i> Consider expanding the scale (e.g., from 3 to 4 points) to accommodate these “border dwellers.”</p>
<ul style="list-style-type: none"> • Are scores determined quantitatively; i.e., by “counting on fingers”? 	<p><i>If so...</i> Substitute qualitative descriptors for numbers so that differences in salient qualities are characterized within the various score points.</p>

Source: Arter, J. and McTighe, J. (2001). *Scoring Rubrics in the Classroom: Using Performance Criteria for Assessing and Improving Student Performance*. Thousand Oaks, CA: Corwin Press

Questions To Ask When Examining Student Work

Use the following questions to guide the examination of student work.

Describe

- *What knowledge and skills are assessed?*
- *What kinds of thinking are required (e.g., recall, interpretation, evaluation)?*
- *Are these the results I (we) expected? Why or why not?*
- *In what areas did the student(s) perform best?*
- *What weaknesses are evident? • What misconceptions are revealed?*
- *Are there any surprises? • What anomalies exist?*
- *Is there evidence of improvement or decline? If so, what caused the changes?*

Evaluate

- *By what criteria am I (are we) evaluating student work?*
- *Are these the most important criteria?*
- *How good is “good enough” (i.e., the performance standard)?*

Interpret

- *What does this work reveal about student learning and performance?*
- *What patterns (e.g., strengths, weaknesses, misconceptions) are evident?*
- *What questions does this work raise?*
- *Is this work consistent with other achievement data?*
- *Are there different possible explanations for these results?*

Identify Improvement Actions

- *What teacher action(s) are needed to improve learning and performance?*
- *What student action(s) are needed to improve learning and performance?*
- *What systemic action(s) at the school/district level are needed to improve learning and performance (e.g., changes in curriculum, schedule, grouping)?*

• Other: _____?

• Other: _____?

Data-Driven Improvement Planning

Based on an analysis of achievement data and student work:

- What *patterns* of weakness are noted?
- What *specific* areas are most in need of improvement?



What *specific* improvement actions will we take?

Data-Driven Improvement Planning

Based on an analysis of achievement data and student work:

- What *patterns* of weakness are noted?
- What *specific* areas are most in need of improvement?

- problem solving and mathematical reasoning are generally weak
- students do not effectively explain their reasoning and their use of strategies
- appropriate mathematical language is not always used
- _____



What *specific* improvement actions will we take?

- Increase our use of “non routine” problems that require mathematical reasoning.
- Explicitly teach (and regularly review) specific problem solving strategies.
- Develop a poster of problem solving strategies and post in each math classroom.
- Increase use of “think alouds” (by teacher & students) to model mathematical reasoning.
- Develop a “word wall” of key mathematical terms and use the terms regularly.
- Revise our problem solving rubric to emphasize explanation & use of mathematical language.

ANNOTATED EXEMPLAR
Persuasive

School is meant to be a place of learning, an opportunity to acquire knowledge and insight, and it was at Greece Olympia High School that I learned this lesson. It was one of those rainy day mornings when little could be heard above the squeak of wet rubber soles against the tile floor of the freshman hallway. I was heading into homeroom early; I thought I'd

The writer engages the reader by establishing a context and using an appropriate tone

be the first to arrive. However, just as I was about to enter the room, I saw that a girl with vibrant brown hair, jeans, and a pink sweater had already gone into the room. Seemingly because her shoes had no texture, with a bottom as smooth as the complexion of her youth, she slipped, hung in the air for a moment, then crashed to the ground. I took a step backward to laugh out in the hall. When I peered back in the room, I expected that after such a fall she would be unable to move. However, she had already leapt to her feet. That's when I noticed her fervent glances. Left and right. Left then right. Her head quickly turned. Satisfied in her anonymity, she slowly, and I believe painfully, walked to her seat.

The writer's use of imagery helps to create a context for the reader.

The writer utilizes vivid and precise language.

The writer varies sentence patterns for effect.

At that moment, I became consciously aware that people, including myself, seem to concern themselves more with the opinions and wants of others than with what they themselves think or desire. This girl had been so worried about what someone else might think that she didn't even stop to catch her breath. It's no wonder that a phrase like, "What will the neighbors think?" sounds cliché. For years people have been interested in owning a better house, buying a faster car and having a more attractive mate. Yet, are these things going to bring self-fulfillment? Is somehow having these items going to impress people, and, if so, why do we care what these people think? We are raised to do just that. From a young age, we are taught to please mostly our parents, then our teachers, coaches, and friends. From the moment we are born, others expect us to behave, think, and value in a certain way, and being the impressionable youths that we are, we usually unwittingly comply.

The writer chooses and employs specific rhetorical devices to support assertions and strengthen persuasiveness of the argument (anecdote) based on the topic, audience and purpose.

The writer uses effective interpretation that offers insights.

Anchoring Performance Assessment Tasks

Anchoring refers to the process of selecting examples of student work/responses to characterize each of the score points on a rubric scale. These examples, known as anchors, provide tangible and specific illustrations of various levels of performance or degrees of proficiency based upon established criteria. Anchors serve an important role in performance assessment by:

- assisting teachers in understanding and consistently applying the scoring criteria when judging student responses;
- providing teachers with student examples for instructional use;
- offering teachers and students clear targets and examples of excellent performance to motivate and guide their efforts; and
- helping students to understand and apply the criteria when evaluating their own work

Models for Anchoring

There are two basic models for anchoring performance tasks. Model 1 is based upon the use of established scoring criteria contained in a scoring tool (rubric, rule, or key). In this model, student responses, products, or performances are evaluated according to the scoring criteria. Then, the scored responses are sorted into groups corresponding with the various score points on the scale (4's, 3's, etc.). Several responses, products, or performances are selected from each group to illustrate the criteria for that score point. These are the anchors.

Model 1 is appropriate when a performance task and accompanying scoring tool(s) have been validated through reviews, field testing, and revision.

Model 2 uses student responses, products, or performances as the basis for identifying or refining the scoring criteria. In this model, student responses are sorted into three (high, medium, low) or four (excellent, good, fair, poor) groups based upon general quality. Each group is then reviewed to determine the distinguishing characteristics of the responses. Specific criteria are then developed for each group, and several responses are selected as anchors to illustrate those criteria.

Model 2 is appropriate when a task has been used for the first time and when there is no scoring rubric or the tool is in draft form.

Task Anchoring Process - Two Models

```
graph TD; A[Task Anchoring Process - Two Models] --> B[Model 1 Based on Scoring Criteria]; A --> C[Model 2 Based on Student Responses];
```

Model 1 Based on Scoring Criteria

- Review the scoring tool to become familiar with the range and criteria for each score point.
- Follow a consensus process to evaluate student responses using the scoring tool.
- Sort the scored responses into groups corresponding with the score points (4's, 3's, etc.).
- Select several responses that best illustrate the distinguishing characteristics for the top score point. These are the anchors.
- Repeat the process for the other score points.

Use Model 1 when...

- the task has been validated through reviews, field testing, and revision,

and

- the scoring tool (rubric, rule, or key) has been validated.

Model 2 Based on Student Responses

- Follow a consensus process to sort student responses into three groups – "high," "medium," or "low" quality .
- Decide on the distinguishing characteristics of the "high quality" responses.
- Use these characteristics to identify the criteria for the top score point of the scoring tool.
- Select several responses that best illustrate the distinguishing characteristics for the top score point. These are the anchors.
- Repeat the process for the other groups of student responses.

Use Model 2 when...

- the task and the scoring tool are being tried for the first time

or

- the scoring tool is in draft form and has not yet been validated.

An Anchoring Procedure – Model 1

This procedure is intended for use in conjunction with an established scoring rubric to identify tangible examples (anchors) to illustrate the different levels of performance specified by the rubric scale.

Have teachers meet in role-alike groups (e.g., grade level or department groups) to evaluate a set of student responses, products, or performances on a common performance assessment task. A group size of 3 or 4 people is recommended. The group uses an established scoring rubric to evaluate student performance according to the following procedure:

1. Collect a full range of student responses for each assessment task. Whenever possible, these responses should be obtained from students of varying achievement levels in different classes.
2. Identify teams with three to four members to work on evaluating and anchoring.
3. Prior to scoring, review each task to become familiar with the identified content standards (or desired understandings) being assessed. Also, review the scoring rubric for each task to become familiar with the range (number of score points) of the scale and the criteria for each score point.
4. Work individually to score designated student responses. Use one of the corners of the Anchoring Form (Figure x.x) to record your score for each task. In other words, each group member privately writes an “H” (for High), an “M” (for Middle), or an “L” (for Low) and folds the corner down so that it can’t be seen. The paper and the accompanying form is then passed to the next rater.
5. Compare the individual scores within the group. Reach consensus through discussion. If necessary, request a “second opinion” from someone not within the group.
6. Once scoring has been completed for a given task, sort the student responses into groups according to their scores; e.g., ones, twos, threes, etc. Then arrange the responses hierarchically within the piles, looking for performance gradations.
7. Examine the responses in each group and look for common features. Select two or three examples of student responses that best illustrate each point on the scoring rubric. These examples will serve as “anchors” for the scoring system.

An Anchoring Procedure – Model 1

(continued)

8. Complete the annotation section on the Anchoring Form for each example selected as an anchor. Be specific, using the language of the rubric to highlight the key features of the response. These annotations should describe why the response received its score so as to assist other teachers in applying the scoring rubric.

Tips/variations for this procedure

- Individuals involved in evaluating and anchoring should be thoroughly familiar with the assessment task(s). Prior to scoring, it is beneficial to work with a partner/team to clarify precisely what students are being asked to do in the task (and its overall purpose) so that the most salient performance features are evaluated.
- It is important to discuss the meaning of each criteria in the scoring rubric so that evaluators will be looking at student performance through the same lens. Also, discuss the differences in the various score points in the scale (i.e., what distinguishes a “3” from a “2”?).
- Scoring and anchoring of performance assessments require the application of human judgment guided by specific criteria. Scoring reliability is strengthened when judgments are reached through a consensus process involving two or more scorers.
- Beware of the tendency to slip into “norm-referenced” evaluation when judging student performances. This can occur when responses are judged according to the best performance in the group rather than against the established scoring criteria.
- Avoid “double jeopardy” scoring. For example, if a student makes a computational error on a mathematics task, don’t let all subsequent responses be penalized because of an initial error, especially when the student demonstrates sound reasoning.
- The Anchoring process described above is appropriate when the performance task and the scoring rubric have been validated through reviews and field tests. The following variation is suggested when the performance task and/or the scoring rubric are being tried for the first time.
 - Follow a consensus process to sort student performances into three groups – “high,” “medium,” or “low” quality.
 - Agree on the distinguishing characteristics of the “high quality” (H) responses. Use these characteristics to identify the criteria for the top score point of the scoring rubric.
 - Select several responses that best illustrate the distinguishing characteristics for the top score point. These are the anchors.
 - Repeat the process for the other groups of student performances to develop and flesh out the rubric descriptors and corresponding anchors for the other score points in the rubric scale.

An Anchoring Procedure – Model 2

The following procedure is designed to guide groups in determining the characteristics of student work/responses responses to performance tasks according the various levels on a rubric and identifying anchor examples. This process also helps teams refine their tasks and associated scoring tools,

1. Each group member randomly selects five student responses to the same task.
2. Each of the four group members begins by reading the first activity (for tasks with multiple activities and scoring tools) or reads the entire task, if the task is to receive a single holistic score.
3. Each person reads the student response on the first paper and globally decides whether the response is closest to High, Middle, or Low quality.
4. In the upper left hand corner, each group member privately writes an “H” (for High), an “M” (for Middle), or an “L” (for Low) and folds the corner down so that it can’t be seen. The paper is then passed to the next rater.
5. Continue with the next paper and the next corner (upper right, etc.) until each paper in the stack has been rated by all.
6. As a group, discuss each paper in turn as the corners are folded open and come to a consensus as to whether papers are “High,” “Middle,” or “Low” quality.
7. Parcel out the "high quality" papers and decide on the distinguishing characteristics of a high quality response, and record them using as much detail as necessary on the sheet entitled, “Characteristics For High Quality Responses.”
8. Use these characteristics to identify the criteria for the top score point of the scoring tool.
9. Revise the scoring tool as necessary to incorporate these criteria.
10. Select several papers that best illustrate the distinguishing characteristics of a high quality response to serve as anchors for the top score point.
11. Repeat the process for the papers in the "middle" and "low" quality groups.

MARYLAND ASSESSMENT CONSORTIUM

Scoring and Anchoring Annotation Form

Title of Task: Take a Hike

Date of Anchoring: 3/29/'04

Annotation:

This is an example of a **4** because:

- computation is completely accurate
- problem-solving strategies are effectively used
- explanation of process is clear and complete with excellent use of mathematical language

Group Members:

Name

School

- | | |
|-------------------|----------------------------|
| 1. Jerry Bruner | Progressive Charter School |
| 2. Matilda Hunter | John B. Goode Elementary |
| 3. Harry Dewey | Kripple Kreek Elementary |

Anchoring Annotation Form

Title of Task: _____

Activity Name or #: _____

Annotation: This is an example of a _____ because:

(score point)

Group Members:

Name _____ LEA

1. _____

2. _____

3. _____

4. _____

5. _____

Group Leader(s):

1. _____

2. _____

A group of 8 people are all going camping for three days and need to carry their own water. They read in a guide book that 12.5 liters are needed for a party of 5 people for 1 day. Based on the guide book, what is the minimum amount of water the 8 people should carry all together? Explain your answer.

A

$$8 = \text{stick figure} \text{ stick figure} \text{ stick figure} \text{ stick figure} \text{ stick figure} \text{ stick figure} \text{ stick figure} \text{ stick figure}$$

$$\begin{array}{r} 12.5 \\ \div 5 \\ \hline 2.5 \end{array} \times 4 = 10 + 50 = 60$$

B

I divided 12.5 liters by 5 people = 2.5 liters/person. I did that so that I could take 2.5 liters X 8 people = 20 liters each day. Now I need to multiply 20 liters/day X 3 days = 60 liters to last the whole camping trip.

C

20 liters for 8 people one day

2.5 for each person

$$\begin{array}{r} 12.5 \\ \div 5 \\ \hline 2.5 \end{array}$$

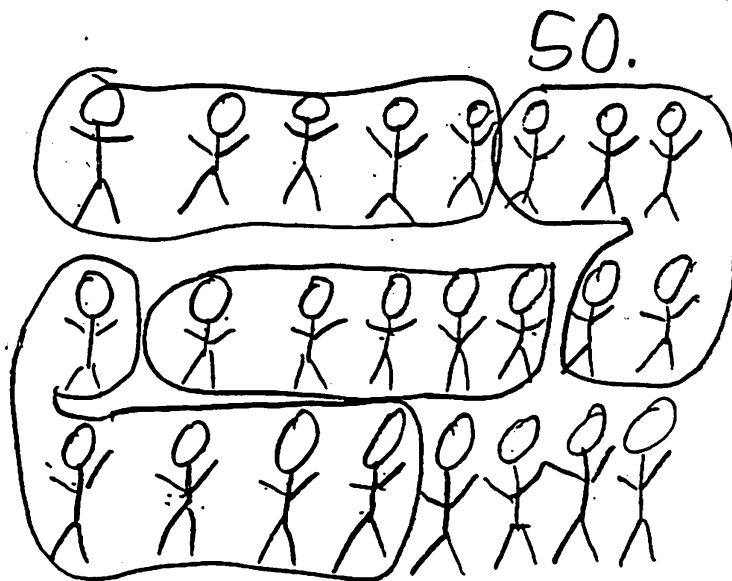
$$\begin{array}{r} 12.5 \\ + 2.5 \\ \hline 15.0 \\ + 2.5 \\ \hline 17.5 \\ + 2.5 \\ \hline 20.0 \end{array}$$

A group of 8 people are all going camping for three days and need to carry their own water. They read in a guide book that 12.5 liters are needed for a party of 5 people for 1 day. Based on the guide book, what is the minimum amount of water the 8 people should carry all together? Explain your answer.

D

They can bring enough water for 5 people for 6 days so 75 liters should be enough

E



F

If 12.5 liters are needed for 9 people for 1 day then 45 liters will be needed for three days for 8 people because for 1 person it was 1.5 which all together equaled 12.5 for 1 day for 8 people to go on a 1 day trip you'd have to bring 65 liters.

A group of 8 people are all going camping for three days and need to carry their own water. They read in a guide book that 12.5 liters are needed for a party of 5 people for 1 day. Based on the guide book, what is the minimum amount of water the 8 people should carry all together? Explain your answer.

G

60 liters. I came up with that by getting how much 1 person needed then how much 8 people needed that was 20 then I multiplied that by three

H

$$\begin{array}{r} 12.5 \\ \div 5 \\ \hline 4 \end{array}$$

$$\begin{array}{r} 12.5 \\ + 2 \\ \hline 24.7 \end{array}$$

$$\begin{array}{r} 24.5 \\ + 4 \\ \hline 28.5 \end{array}$$

28.5

I

If 12.5 liters are needed for 9 people for 1 day then 45 liters will be needed for three days for 8 people because for 1 person it was 1.5 which all together equaled 12.5 for 1 day for 8 people to go on a 1 day trip you'd have to bring 65 liters.

J

They can just go to 7-11 and each person buys a 6 pack of bottled water to bring.

Scoring and Annotation Form

	High (3)	Middle (2)	Low (1)
R E A S O N S			