Quality Assessment Items

Tom Schimmer
Assessment Author and Associate
Email: tschimmer@live.ca
Twitter: @TomSchimmer #GFIO #atAssess
Facebook: Schimmer Education
www.allthingsassessment.info
“Assessment methods are not interchangeable. To ensure accurate assessment results, the overriding criterion for selection of method is consideration of the type of learning targets to be assessed: Some assessment methods are better matches for certain types of learning targets than others.”

Chappuis & Stiggins, 2019

### Standards and Cognitive Complexity

**Deconstructing a Standard (Erkens, 2016)**

1. **CIRCLE** all verbs.
2. **UNDERLINE** anything that requires direct instruction.
3. **STAR** any skills that can only be scored in the *doing* and not the *done* phases (intentionally watch and listen during performance).
4. **BOX** any components that need to be part of the final product.

### Assessment Methods

<table>
<thead>
<tr>
<th></th>
<th>Advantage</th>
<th>Challenge</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Performance Assessment</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Constructed Response</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Selected Response</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Learning Progression vs Progressions of Quality

- **Learning Progression**: An instructional sequence that leads students through the taxonomy of learning; primarily focused on the formative purpose and when students are ready for an increase in complexity.

- **Progression of Quality**: A scale that describes the simple to sophisticated demonstrations of the intended learning outcome or goal; primarily focused on the summative purpose and the verification that learning has occurred (and to what degree).

“A learning progression is a model of successfully more sophisticated ways of thinking about a topic typically demonstrated by children as they learn, from naïve to expert.”

—National Research Council (2007)

“Learning progressions can provide teachers with a **blueprint for instruction & assessment** because they represent a goal for summative assessment, indicate a sequence of activities for instruction, and can inform the design of formative assessment processes that provide indicators of students’ understanding. The value of learning progressions for CA lies in the information they provide about **what to assess and when to assess it**.”

—Heidi Andrade (2013)

**Developing Learning Progressions**
Developing Success Criteria

Build a Rubric?

- “Is this demonstration scalable?”
- YES: Build a rubric focused on a progression of quality.
- NO (Binary): Build a rubric focused on the level of consistency.

Why Rubrics?

- Teachers must be clear on what to look for in student demonstrations.
- Students must be clear on what to look for in their (or others’) demonstrations.

“The biggest mistake teachers make when they use rubrics with performance assessment is that they focus on the task, the product, and not the learning outcome or proficiency the task is supposed to get students to demonstrate.”

-Susan Brookhart, (2013)

Rubric Types

<table>
<thead>
<tr>
<th>Type</th>
<th>Description</th>
<th>Advantage</th>
<th>Disadvantage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Analytic</td>
<td>Breaks down the criteria into specific aspects of quality</td>
<td>Specificity as to what is strong and what needs strengthening; excellent for formative assessment.</td>
<td>Challenging to create. Can be cumbersome for summative assessment, especially when an overall level of quality is being determined/recorded.</td>
</tr>
<tr>
<td>Holistic</td>
<td>Provides a much broader, overall description of quality along several levels (i.e. 3-5)</td>
<td>Reliability with scoring inferences is much more readily attained; excellent for summative assessment.</td>
<td>The lack of detail makes them a challenge to use for instruction and feedback purposes.</td>
</tr>
<tr>
<td>Single-Point</td>
<td>Breaks down the criteria into specific aspects of quality, but only describes the highest level of proficiency.</td>
<td>The space beside each specific aspect is used for personalized comments on what is strong and what needs strengthening.</td>
<td>Significantly more time consuming to complete, which could increase the turnaround time for feedback.</td>
</tr>
</tbody>
</table>

Adapted from: Balch, Blanck, & Balch, 2016
Selected Response Items

1. Use new/original material to elicit higher level thinking skills.
2. Each question should be self-contained.
3. List answer choices vertically.
4. Write questions at the appropriate reading complexity.
5. Minimize the amount of reading in each item.
6. Ensure that one answer is clearly correct.
7. Make sure all distractors are plausible.
8. List options in numerical or logical order.
9. Ensure that no clues embedded within the question.

<table>
<thead>
<tr>
<th>Guideline in Question</th>
<th>Yes</th>
<th>No</th>
<th>Why?</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) Question stems should be written as complete questions.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2) Question stems should be stated positively.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3) Avoid using ALL OF THE ABOVE and NONE OF THE ABOVE</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4) Keep answer choices equal in length</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5) Three answer choices are sufficient?</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

“...items written in either open or closed form should contain a single problem or restricted problem scope, and the item stem, in either format, should contain all necessary information to select the best answer.”


**Gross, 1994**

- Logical rather than empirical guidelines.
- Avoid using any features that make discriminating between test takers more difficult.
- Correct answers can be obtained with misinformation.
Frary, 1991

- When NOTA is the correct answer it can (potentially) motivate students to consider every option.
- Recognizing NOTA may actually reflect greater understanding of the material vs. finding one right answer.
- NOT when which is best type questions are asked.
- NOT when at least one option is illogical.

Haladyna & Downing, 1991

- The quality of the options is significantly more important than the quantity.
- Students tend to make educated guesses by quickly eliminating the least plausible option.
- Practically speaking, writing a 4th (or 5th) quality item is challenging.
- Also, more efficient for students and teachers.

### Constructed Response Items

“Of all the issues raised by the use of test response formats, anticipation effects may be the most important in the long run. Although it is commonly believed that expectations concerning different response formats lead to different learning outcomes, this connection is still not well-established and described in the research.”


**Extended Written Response (Chappuis & Stiggins, 2019)**

1. **Set the Context:**
   - *We have been studying the inheritance and variation of traits through heredity*

2. **Tell what to describe or explain.**
   - *Based on your understanding, make and defend the claim that inherited genetic traits can and do vary.*

3. **Point the way to an appropriate response.**
   - *Be specific in your descriptions and be sure to include how variations can occur through meiosis, how errors in replication occur, and the source of various mutations.*

*We have been studying the inheritance and variation of traits through heredity. Based on your understanding, make and defend the claim that inherited genetic traits can and do vary. Be specific in your descriptions and be sure to include how variations can occur through meiosis, how errors in replication occur, and the source of various mutations.*
Constructed Response (Long Answer)

- Is there a more effective/efficient way for you to access the learning information you’re seeking?
- Are the students familiar with the rubric/success criteria and how it will be applied?
  - SAME RUBRIC? Are there any shifts in emphasis that students need to be familiar with?
  - DIFFERENT RUBRIC: Have you identified the differences between the “familiar” and “unfamiliar” rubrics?
- Is my question clear and thorough enough to avoid any confusion about what content to include?

Performance Assessment

“Performance assessments are demonstrations of mastery that emulate the context or conditions in which the intended knowledge or skills are actually applied.”

-AERA, APA, & NCME, 1999

Why Performance Assessment (Linn, 1993)

1. Allow for demonstrations of important and meaningful learning targets that cannot be easily assessed with other formats.
2. They serve as exemplars of tasks that stimulate and enrich learning rather than just serve as indicators of learning.
3. They help shape sound instructional practices by modeling to teachers what is important to teach and to students what is important to learn.
Intent (Lane, 2010) | Clarity (Mislevy, Steinberg, & Almond, 2003)
--- | ---
- **Generalization**? Then adequate sampling across the domain or discipline is required. | - We must be clear on both the **content** and the **cognitive processes** that are being assessed.
- **Finite performance**? Then more specific tasks & criteria are more appropriate. | - What **performances** will reveal a level of proficiency with that content and/or cognitive processes.
- **Formative**? Then a much narrower scope is more fitting. | - What **tasks** will most likely lead to those performances.

**Using Rubrics**

1. Thoroughly explain the rubric AND/OR co-construct the rubric/success criteria with students.
2. Allow students to practice using the rubric (exemplars) prior to it being applied to their work.
3. Provide clear direction on meaning, language, and phrasing if the rubric is to be used for peer assessment activities (self-assessment).
4. Make sure the rubric is directly connected to the specific standard(s) being assessed.
5. Be thoughtful about using generic rubrics. (+) Clear & Consistent (-) Potentially an awkward fit for specific assignments.

**Rubric Types**

<table>
<thead>
<tr>
<th>Type</th>
<th>Description</th>
<th>Advantage</th>
<th>Disadvantage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Analytic</td>
<td>Breaks down the criteria into specific aspects of quality</td>
<td>Specificity as to what is strong and what needs strengthening; excellent for formative assessment.</td>
<td>Challenging to create. Can be cumbersome for summative assessment, especially when an overall level of quality is being determined/recorded.</td>
</tr>
<tr>
<td>Holistic</td>
<td>Provides a much broader, overall description of quality along several levels (i.e.3-5)</td>
<td>Reliability with scoring inferences is much more readily attained; excellent for summative assessment.</td>
<td>The lack of detail makes them a challenge to use for instruction and feedback purposes.</td>
</tr>
<tr>
<td>Single-Point</td>
<td>Breaks down the criteria into specific aspects of quality, but only describes the highest level of proficiency.</td>
<td>The space beside each specific aspect is used for personalized comments on what is strong and what needs strengthening.</td>
<td>Significantly more time consuming to complete, which could increase the turnaround time for feedback.</td>
</tr>
</tbody>
</table>

*Adapted from: Balch, Blanck, & Balch, 2016*
“The biggest mistake teachers make when they use rubrics with performance assessment is that they focus on the task, the product, and not the learning outcome or proficiency the task is supposed to get students to demonstrate.”

-Susan Brookhart

_How to Create and Use Rubrics for Formative Assessment, p. 15_

---

### Subject-Specific Guidelines

<table>
<thead>
<tr>
<th>English Language Arts</th>
<th>Mathematics</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Items that are text-dependent and measure more than literal comprehension.</td>
<td>1. Targets core disciplinary processes and ideas.</td>
</tr>
<tr>
<td>2. Items that measure cognitive and metacognitive competencies for reading literacy.</td>
<td>2. Allow for multiple entry points.</td>
</tr>
<tr>
<td>3. Items that require evidence-based writing.</td>
<td>3. Allows for multiple solution strategies.</td>
</tr>
<tr>
<td>4. Items that call for authentic disciplinary uses of literacy.</td>
<td>4. Considerate presentation of item.</td>
</tr>
</tbody>
</table>

### Science

1. Assessments should be aligned with and integrate multiple learning dimensions of science and engineering.
2. Assessments should focus on the big ideas in science.
3. Assessments should address the full range of science and engineering practices.
4. Assessments should require students to demonstrate their reasoning and problem-solving skills.

### History

1. Does the item prioritize domain-specific skills and knowledge?
2. Does the item represent a facsimile of disciplinary work, both in the materials used and what students are asked to do for the item?
3. Are the materials used in the item carefully selected and prepared to minimize confounding factors such as reading ability and necessary background knowledge?
4. Are there multiple pathways to a correct answer or multiple correct responses?
5. Is the item “balanced” in construction?

*Source: Stanford Center for Assessment, Learning, and Equity (SCALE)*
References


