

Keys to Quality Assessments

Before using any assessment, the user must ensure its quality. Any standardized test, state- or district-developed assessment, or classroom assessment must be evaluated through the careful application of specific quality control standards. Good consumers ask tough questions about quality. We present here the questions that an assessment user or author should ask in conducting such an analysis of quality. These questions are framed as a rubric that covers each of the five Keys to Quality Assessment:

Key 1—*Clear Purpose*: A sound assessment serves clearly articulated and appropriate purposes—why is the assessment taking place, who will use the information, and what will the information be used for?

Key 2—*Clear Targets*: A sound assessment arises from clear and appropriate student learning target(s)—achievement expectations are clearly and completely defined and reflect the best current understanding of the field.

Key 3—*Sound Assessment Design*: A sound assessment is designed with learning targets and purposes in mind. It uses an appropriate assessment method; has well-written items, tasks, and rubrics; samples student achievement in such a way to make appropriate inferences about student learning; and avoids potential sources of bias that could distort results.

Key 4—*Good Communication*: Communication is planned as part of the assessment to serve the needs of users.

Key 5—*Student Involvement*: A sound assessment involves students in self-assessment, recording keeping, and/or communication.

Prior to the rubrics themselves, there is a summary sheet titled “Questions to Ask of Classroom Assessments” that serves as a bridge between the definitions of each key presented above and the much more detailed rubrics themselves. Following the detailed rubrics is a short version in checklist form. Users often find both the shorthand and the full-version rubrics useful.

Rationale for the Scale

We have three-point rubrics—an assessment can be “fast tracked,” “on track” (but needs work), or “side tracked” for each of the five Keys to Quality Assessment. The rubrics can easily be converted into a five-point scale. Think of a “4” as having some qualities of “fast tracked” and some of “on track.” Likewise, a “2” can be thought of having some qualities of “on track” and some of “side tracked.”

Source: Adapted from *Practice with Student-Involved Classroom Assessment* (pp. 349–358), by J. A. Arter & K. U. Busick, 2001, Portland, OR: Assessment Training Institute. Adapted by permission.

When First Reading Each Rubric

Read the rubrics in this manner: First read the “fast tracked” level. Highlight words and phrases that jump out at you as truly defining high quality. Then read “side tracked.” Again highlight words and phrases that serve to define weak quality for you. Finally, read “on track.”

Scoring

First, please note that *scoring* does not mean giving the assessment a judgmental grade. We use the term in a formative sense, not a summative one. The word is shorthand for “analyzing an assessment for quality so that we get good at recognizing features that are productive and counterproductive.”

A couple of procedural thoughts on scoring: When you first look at an assessment that you want to analyze for quality, decide if it is stronger or weaker on the key (trait) you are considering. If it is stronger, begin reading at the “fast tracked” level. If the assessment is not that strong, jump to the “on track” descriptions. If the assessment is stronger than “on track” but not quite “fast tracked,” consider adopting a five-point scale and giving it a “4.”

Likewise, if your first look at an assessment indicates that it is weak on the key (trait) under consideration, begin reading at the “side tracked” level. If it is stronger than that, jump to the “on track” level. If the assessment is stronger than “on track” but not quite “fast tracked,” consider adopting a five-point scale and giving it a “2.”

Sometimes it is helpful to actually highlight words and phrases on the scoring guide that describe what you are seeing in the assessment under review. This helps you focus on where the constellation of features falls.

A Note of Caution

Do not think of these rubrics as checklists—it is not true that everything under “fast tracked” has to be present for an assessment to get a high score. Rather, the statements in each level of the rubrics represent the types of things characteristic of an assessment at each level of quality. Use the rubrics by finding the descriptors that most match the assessment you are reviewing.

Questions to Ask of Classroom Assessments

Key 1: Clear Purposes

Do the assessment authors have a clear picture of how the assessment results will be used and by whom?

- Is it clear who will use the results and how they will be used? Is the distinction between assessment *for* and *of* learning clear? (See CASL Chapter 2.)
- How do the purposes for this assessment fit into a bigger plan that addresses both assessment *for* and *of* learning over time? (See CASL Chapters 4 and 9.)

Key 2: Clear Targets

Do the assessment authors have a clear picture of what they are trying to measure? (See CASL Chapter 3.)

- Are the student learning targets stated and easy to find?
- Are the student learning targets focused—there are not too many?
- Are they clear? Would teachers agree on what they mean?
- Are they appropriate? Do they represent the heart of the discipline and are they worth the instructional and assessment time devoted to them? Is there a clear connection to standards?
- Do the learning targets assessed represent what was taught?
- Do the stated learning targets fit into a bigger plan to cover all important learning targets over time? Do they fit into a bigger plan across grade levels—previous and next learning—in a continuous-progress curriculum?

Key 3: Good Assessment Design

Have the assessment developers translated the learning targets into high-quality assessments that will yield accurate results?

Choosing the Best Assessment Method

Has the assessment method been consciously chosen to fit the learning targets to be assessed and the purpose for the assessment? (See CASL Chapter 4.)

- What types of learning targets are being assessed?
- What are the assessment methods used?

- Are the assessment methods best for the learning targets being assessed? (*Best* is a balance between *most accurate* and *practical*, taking into account the intended use of the information.)

Finding and Devising Quality Tasks and Questions

Are the assessment questions or exercises written well—not confusing, the answer to one question does not give away the answer to another, etc.? (See CASL Chapters 5–8.)

If a performance assessment, is the scoring guide (rubric) clear and does it cover the most important aspects of what defines quality? Does it match the learning targets it is intended to assess? (See CASL Chapter 7.)

Sampling

Does the assessment gather enough information for you to generalize about level of student learning on the target? If not, is this assessment part of a bigger plan to gather sufficient information across time/assessments? (See CASL Chapter 4.)

Avoiding Potential Sources of Bias and Distortion

Do you notice anything in the assessment or way the assessment is carried out that might keep students from adequately demonstrating what they know and can do? This includes unclear targets, unclear purposes, not the best assessment method, problems with tasks and rubrics, inadequate sampling. It also includes problems with context factors and students. (See Chapter 4.)

Key 4: Good Communication

Have the assessment developers planned for adequately managing information from the assessment and reporting it in ways that will meet users' needs? (See CASL Chapter 9.)

- Has communication been planned as part of the assessment?
- Do teachers record assessment information accurately over time and appropriately combine it for reporting?
- Will the users of the results understand them and find them useful?

Key 5: Student Involvement

Are students involved in the assessment? (See CASL Chapters 2, 3, 4, 5, 6, 7, 8, 11, and 12.)

- Would student-involvement components be useful in this case? If so, are they present? This might include making learning targets clear to students, providing descriptive feedback to students, involving students in self-assessment, tracking progress and setting goals, and involving students in communicating about their own learning.
- How does the student involvement in this assessment reflect a bigger plan for involving students in their own assessment?

Assessment Quality Rubrics

| Key/Trait | Fast Tracked = 5 | On Track = 3 | Side Tracked = 1 |
|---|--|--|--|
| <p>Key 1, Clear Purposes: The assessment serves a clear and appropriate purpose.</p> | <p>a. It is clear who the intended users and uses are for the assessment and it is clear how these purposes fit into a bigger plan that addresses both assessment for and of learning over time. b. Users and uses are focused—there are not too many—and purposes are appropriate (e.g., informing students, grading, informing others, planning instruction, evaluating instruction).</p> | <p>a. Purposes for the assessment and how it fits into a comprehensive plan are not stated; the reviewer is able to infer them. b. Users and uses are stated, but there is a question about appropriateness; or, the author does not seem to understand the importance of stating them.</p> | <p>a. No purposes are stated; it is not clear why the assessment is being given. Therefore, it is difficult to tell how this assessment fits into a bigger plan to cover assessment for and of learning over time. b. There are too many users and uses; it would be impossible to satisfy all the stated purposes in a single assessment; or, the purposes are inappropriate; you might ask, “Why are they generating information for this purpose?”</p> |
| <p>Key 2, Clear Targets: The assessment reflects valued student learning targets.</p> | <p>a. Targets are stated, clear, focused, and easy to find. b. Targets are important—worth the assessment time devoted to them. Targets are clearly related to state and district content standards, or target descriptions and definitions reflect best thinking in the field. c. It is clear that learning targets came first, then instruction and assessments were designed to track progress toward and help attain the targets. d. There is an appropriate mix of targets and/or there is evidence of long-term thinking—how targets in the current assessment fit with plans for assessing all important targets over the year.</p> | <p>a. Targets are listed, but they might be stated differently in different places or require some work to find. b. Some targets are important; they are worth the assessment time devoted to them, some are not; or, the connection to state or district standards is not entirely clear; or, some of the targets represent best thinking in the field, others do not. c. Targets seem to be retrofitted to an already existing assessment; as a result, you might feel somewhat dissatisfied that the assessment is well thought out. d. Long-term planning (how targets in the current assessment fit with plans for assessing all important targets over time) must be inferred.</p> | <p>a. Stated targets are broad and/or vague; there is little attempt at clarification; or, no targets are stated; or, statements of targets ramble; or the author lists one and later lists others; or, targets have to be inferred from the assessment itself; or, too many are listed. b. If targets are stated, they seem trivial; why spend time assessing this? There is little apparent connection to district or state standards, or the description of targets does not reflect best thinking in the field. c. Activities for students are planned without taking learning targets into account. d. There is a poor mix of targets: the author might have chosen only the easiest ones to assess; or, there is little evidence of how the targets in the current assessment fit into the overall plan for assessing all important targets over the year.</p> |

Source: Adapted from *Practice with Student-Involved Classroom Assessment* (pp. 349–358), by J. A. Arter & K. U. Busick, 2001, Portland, OR: Assessment Training Institute. Adapted by permission.

| Key/Trait | Fast Tracked = 5 | On Track = 3 | Side Tracked = 1 |
|--|---|--|---|
| <p>Key 3, Sound Assessment Design: Learning targets are translated into assessments that yield accurate results.</p> | <p>Choosing the Best Method a. The method selected is capable of providing information about the learning target. Writing Questions b. Instructions are clear. c. Tasks and exercises are feasible—it is possible for well-prepared students to complete them successfully. d. Selected response items adhere to standards of quality. e. When the assessment is a performance assessment or extended written response, performance criteria are clear, well-defined, and cover the most salient features of a performance. Sampling f. There are not too many nor too few tasks or exercises, but just enough to get a stable estimate of learning, and the tasks cover the learning target(s) (domain) well. If this assessment consists of a single task, it is clear how it is part of a bigger plan to gather sufficient information across time/assessments. g. There is a clear match between stated learning targets and items on the assessment. Sources of Bias h. Accommodations for diverse student characteristics are made and such accommodations do not affect the accuracy of the final judgment of student learning. i. Cultural or gender differences will not interfere with a student's ability to accurately demonstrate their learning. j. Possible sources of bias and distortion are described or acknowledged; none are too severe; caveats on use are given.</p> | <p>Choosing the Best Method a. Target-method match may be satisfactory for part of the assessment; some of the methods might be the best choice. Writing Questions b. Although some tasks or exercises have features that are vague or confusing, they are generally sound. c. Some tasks and exercises are feasible, some are not. d. Some selected response questions adhere to standards of quality, some do not. e. When the assessment is a performance assessment, although performance criteria may be a little vague or confusing, they are generally sound and just require some adjustments or rewording. Sampling f. In some cases there are too many or too few tasks or exercises to get an efficient estimate of student learning; or, some outcomes are sampled well, some are not. g. The assessment reflects most, but not all, of the stated learning targets; or, the assessment reflects learning of some targets not stated. Sources of Bias h. There are some accommodations for diverse student characteristics, and/or accommodations might sometimes affect final judgment of student learning. i. The assessment might work well for one group of students, but might need to be reworked for use with other groups. j. Although the authors are aware that bias and distortion can be a problem, they do not completely address potential problems in the assessment; or, information about bias and distortion has to be inferred or searched for.</p> | <p>Choosing the Best Method a. The method does not seem capable of doing the job—you might ask, "Why did they assess the target <i>that way</i>?" Writing Questions b. Tasks or exercises are vague or confusing, and it is difficult to see how they might be fixed. c. It would be hard to do the tasks or exercises successfully; e.g., a task requires specialized equipment that might not be equally available to all students. d. Selected response questions do not adhere to standards of quality. e. When the assessment is a performance assessment, performance criteria do not cover important elements of performance, are vague or confusing, do not match the stated target(s), or are missing entirely. Sampling f. There are not enough tasks to draw the desired conclusion; e.g., the author draws a conclusion about student ability to think critically on the basis of a single passage. g. The assessment reflects only a few of the stated learning targets; or, the assessment covers lots of targets not stated. Sources of Bias h. There are no accommodations for diverse student learning characteristics—or accommodations are likely to give an inaccurate picture of student learning. i. Tasks and procedures might unfairly cause different groups of students to do poorly even when their skills and knowledge are the same. j. It is easy to identify several severe potential sources of bias and mismeasurement. The authors seem unaware of possible sources of bias and mismeasurement.</p> |

| Key/Trait | Fast Tracked = 5 | On Track = 3 | Side Tracked = 1 |
|--|--|---|---|
| <p>Key 4, Good Communication: Assessment results are managed well and communicated effectively.</p> | <p>Managing Information</p> <p>a. Teachers record results carefully over time, keep it confidential, and combine it into an accurate summary of student achievement on intended learning targets. (See the rubric for grading in Chapter 10.)</p> <p>Communication</p> <p>b. Communication is planned as part of the assessment. If there is no communication component, there is a good reason, as when teachers collect information solely for their own use in planning.</p> <p>The authors have anticipated the needs of the intended users:</p> <ul style="list-style-type: none"> • Timing—Information is provided when needed. • Understandability—The user(s) understand the message; this understanding has been checked. • Type of information—Usually descriptive, but sometimes judgmental depending on whether the use is an assessment <i>for</i> or <i>of</i> learning. • Implications for future student instruction are stated. • Possible unintended negative side effects have been anticipated and avoided. <p>For example:</p> <ul style="list-style-type: none"> • If students are users of the communication, the teacher often provides evaluative feedback rather than a single judgmental evaluation; or, leads students to self-reflect; or, helps students track their progress and set goals for their next steps in learning; or, reviews the assessment with students in small groups and encourages correction, discussion, and further study. • If parents are the recipients of the communication, there is a mechanism for dialogue on next steps; or, parents are given suggestions on how to help their children with the next step in learning; or, students discuss their learning with parents. | <p>Managing Information</p> <p>a. Recorded information is a mixture of traditional and standards-based recording and summarizing. Standards-based information has to be searched for or inferred. (See the rubric for grading in Chapter 10.)</p> <p>Communication</p> <p>b. The communication seems partly to match the communication needs of users, but it still needs work on understandability, timing, descriptiveness, potential for further learning, or possible negative side effects.</p> <p>For example:</p> <ul style="list-style-type: none"> • If students are recipients of the communication, the teacher merely goes over the assignment with students in the whole group with no targeted reteaching. • If parents are the recipients of the communication, although the communication is understandable, it is limited to sending home papers and getting a signature; or, communication is limited to parent ability to get online at any point in time to review the teacher's gradebook. | <p>Managing Information</p> <p>a. Assessment information is not recorded or summarized over time to facilitate standards-based reporting; confidentiality is compromised; the final grade is not an accurate measure of student achievement on intended learning targets. (See the rubric for grading in Chapter 10.)</p> <p>Communication</p> <p>b. The authors have not considered communication at all.</p> <p>Reporting mechanisms do not seem to fit the needs of the users:</p> <ul style="list-style-type: none"> • The stated users might not understand the information. • Information is not timely. • The only information is evaluative when descriptive feedback would serve the purpose better. • Connection to further student learning is not indicated. • Presentation might have negative side effects on the recipient. <p>For example:</p> <ul style="list-style-type: none"> • If the users of the communication are students, they do not receive descriptive feedback, only evaluative feedback such as grades; or, students merely correct each other's papers and call out scores in class. • If parents are the users of the information, parents would not understand the information sent; or, the communication is worded in such a way that parents might be offended. |

| Key/Trait | Fast Tracked = 5 | On Track = 3 | Side Tracked = 1 |
|---|---|---|--|
| <p>Key 5, Student Involvement: Students are involved in assessment.</p> | <p>a. The authors have considered how the assessment results, procedures, and/or materials can be used to do the following:</p> <ul style="list-style-type: none"> • Promote instructional uses. • Help students understand the nature of the learning targets they are to hit through such things as practice with criteria and rubrics, student-friendly versions of rubrics, student development of assessments, or student cross-referencing of assessment questions to goals. • Assist students with self-assessment, tracking learning, and communicating about their learning. <p>b. Student involvement is meaningful—not an add-on that appears to be an afterthought with no stated connection to promoting learning.</p> <p>c. If there exists no student involvement, it is reasonable that there should not be, and it is clear that other assessments are part of a student-involvement plan that unfolds over time.</p> <p>d. There are likely to be positive consequences for students: increased interest in the topic, increased motivation to learn, and increased ability to take control of learning.</p> | <p>a. Student uses are scattered throughout, but have to be searched for or inferred.</p> <p>b. Some student-involvement procedures are meaningful, some are not.</p> <p>c. Although there is some student involvement, the rationale for such use is not stated or related to a larger plan over time.</p> <p>d. The assessment will have more positive than negative effects on student learning.</p> | <p>a. There is no student involvement. Assessment results, procedures, and/or materials are not used to help students understand the learning goals they are to reach.</p> <p>b. Student-involved uses, if present, appear to have been an afterthought, and suggestions do not promote meaningful student involvement; there is no stated connection to increased motivation or learning.</p> <p>c. It is not clear how this assessment is part of a student-involvement plan that unfolds over time.</p> <p>d. The assessment might have negative effects on students' learning; e.g., embarrassment, judging themselves as failures, or turning students off to learning.</p> |