



What makes a strong assessment item?

Scoring Information

An answer key, scoring guide, or rubric should be provided with each item. An answer key for a selected-response item indicates the correct answer, which is assigned a specific number of points.

Scoring guides for a variety of constructed-response item types may include rubrics and sample student responses. A rubric assigns point values to specific components and expectations for the responses to an item. Using rubrics helps ensure that scoring is consistent from scorer to scorer, so that resulting student scores are fair. Rubrics save time in scoring and help educators identify areas where students show understanding of a standard or topic and where they may be struggling. Sharing rubrics with students also helps clarify expectations and promote positive teacher-student interaction to support learning.

Distractors and Distractor Rationales

“Distractors” are the incorrect response choices for a multiple-choice or selected-response item. Good distractors are plausible and based on common misinterpretations or errors. Poor distractors might offer a clue to the correct answer or easily identify that option as an incorrect answer.

“Distractor rationales” indicate which option is the correct answer (the “key”) and identify misconceptions or clarify why the distractors are incorrect. Effective assessments include items with distractors and distractor rationales. They provide insight into student understanding and help you focus on areas of weakness in student learning.

Item Type

All item types—multiple-choice, constructed-response, extended-response, technology-enhanced, performance tasks—elicit different levels of responses from students.

Item rigor is an important element in assuring that assessments deliver what they promise. Item types provide varying levels of cognitive complexity. When determining which item type to use, you should factor in the method of scoring and amount of time allotted for test taking. No matter the item type, items based on evidence-centered design principles define the outcomes and evidence that lead to the most valid measurement. These items elicit meaningful demonstrations of your students’ knowledge and abilities.

Bias and Sensitivity

“Bias” refers to an unintended advantage or disadvantage to some group or groups of students based on construct-irrelevant features that can affect students’ responses to an item. An item may be biased if its content matter caters to or assumes familiarity with a particular experience or environment not equally familiar to all groups. A good item is free of bias.

“Sensitivity” refers to the awareness that certain topics or objects have different perceptions and value in particular cultures. Sensitivity in item content means that items are free of references that might be offensive to a particular culture or group, and the content is acceptable to a diverse audience.

High-quality items include multiple reviews for bias and sensitivity during the development process—leading to more valid and reliable results.

Alignment to Standards

Well-written items measure what is intended by a particular standard. Student responses to well-aligned items provide a valid basis for inferences about attainment of the standard. In addition to review by experts, one way to determine the accuracy of an item’s alignment (and the item’s validity) is to try it out in a field test to see if the student responses show the expected evidence.

Alignment is twofold: First, items must align to your standards and measure what is intended—and second, items addressing a range of standards should be included in a general assessment.

Authenticity

“Authenticity” refers both to the source of a reading passage or other stimulus and to the context of an item. Many assessment programs require that reading passages and other stimuli be authentic—that is, drawn from actual published work—and of a nature and time, current or historic, that would be familiar to the student.

Authentic items are those based on real-world situations and contexts that students recognize and relate to—minimizing opportunities for misunderstanding because they allow the student to connect to an item through personal experience.

Validity and Reliability

“Validity” of an item means the extent to which an item demonstrably measures what it intends to measure; it aligns with the standard and elicits meaningful evidence.

“Reliability” means the extent to which an item or assessment would produce similar results if re-administered.

Valid items and reliable assessments yield data that support your decision making, to inform instruction and measure student growth.