

# REFLECTING ON ASSESSMENT DESIGN

## INTRODUCTION AND PURPOSE

By the end of this module, you should be able to use the assessment checklist to help determine whether an assessment you design appropriately addresses the five elements of assessment design featured in this series of modules.

## KEY CONCEPTS

### The Assessment Checklist

You can use the “Assessment Checklist” to reflect on the design of your assessment. The checklist includes five sets of questions, one for each element of assessment design.



### Alignment

To check for alignment, you can ask whether each assessment item is aligned with the standard you intend to measure and teach in the classroom. If an assessment item is not aligned with the standard you intend to measure, it may measure something else, which can mask what your students actually know and can do. For example, in the module about alignment, we examine the sample item “What is  $12 \div 5$ ?” While this item could be fine in another context, we find that its content is not well aligned with the relevant Tennessee mathematics standard, which states that students should be able to solve division problems using whole numbers. This problem is beyond the scope of the standard because it involves decimals.<sup>1</sup>

### Rigor

To check for an appropriate level of rigor, you can ask whether the level of rigor of each assessment item matches the cognitive complexity you intend to measure and whether the assessment measures a range of student thinking and understanding so that it measures what all students know and can do. If your assessment includes only items that require lower-order thinking, and the skill or skills you intend to measure require lower- and higher-order thinking, your assessment results could indicate that students have mastered the standard or standards when they may not have. For example, the module about rigor includes a multiple-choice item, “Which of the following words is an antonym of ‘tense’?” As we indicated in that module, this item is not well aligned with the rigor of the relevant skill because it merely asks students to *identify* an antonym from a list, when the skill in the standard expects them to be able to *use* the relationship between antonyms to better understand each of the words, which is a higher-level skill.<sup>2</sup>

### Precision

To determine whether your assessment is as precise as possible, you can ask questions to determine whether your assessment items are accurate and clear. The questions you ask will vary depending on the type of assessment items in your assessment. For example, if your assessment includes a number of multiple-choice items, a lack of precision could indicate that students have mastered a standard or standards when they may not have. The right answers may be obvious or easier to guess because your distractors are implausible or because one answer choice is much longer and more detailed than the others.

### Bias

To check for bias, you can ask whether you ensured that each assessment item does not provide an advantage or disadvantage to any group of students because of their personal characteristics, such as race, gender, socioeconomic status or religion. Remember that even though bias results from *personal* characteristics, we detect bias at the *group* level, not at the *individual* level. This distinction is important because an assessment can provide an advantage or disadvantage to individual students without being biased.

### Scoring

To determine whether your assessment has an appropriate scoring strategy, you can ask whether you developed an appropriate scoring tool or tools, and whether your tool or tools are well designed. For example, imagine that you administered a grade-wide essay and that several teachers on your grade-level team scored the students’ assessments. Did you and the other teachers use a common rubric? If so, did you engage in a norming activity in which all raters practiced applying the rubric to the same small number of papers, and then discussed inconsistent ratings to calibrate expectations across all those who would be grading papers? Did your team select student exemplars that are representative of each possible performance level for all raters to use as guidance? If you do not employ techniques like these, each team member will likely apply different standards or look at or emphasize different components of writing. Having a plan to score consistently across classrooms if the assessment is grade wide is a strong assessment practice.

