Four Steps to Creating Valid and Reliable Common Assessments

By Thomas W. Many

How many times have you heard colleagues ask in exasperation, "How am I supposed to write assessments that are valid and reliable? I'm not an assessment expert!"

Many teachers feel they lack the time and expertise they need to develop high-quality assessments. Instead, they often rely on publisher-created assessment items that rob them of one of the best job-embedded professional development opportunities available: authoring their own locally developed common assessments.

> "But when our purpose is a quick determination of the extent to which students understands skills and concepts, and the equally important purpose of adjusting teaching strategies to help students who have not yet mastered those skills and concepts, then practical utility takes precedence over psychometric perfection."
>
> —Douglas Reeves

Best practice calls for teacher teams to use common assessments as a tool to understand how students are progressing through the curriculum. This process transforms the classroom from a place where teachers "teach, test, and hope for the best" to one where teams constantly ask the question, "What did our students learn as a result of our teaching?" Working together, teachers use assessment results to improve instruction and help students deepen their understanding—all of which lead to higher levels of student learning.

How do teachers create these common assessments? Following are four intentional steps teams can take to create valid and reliable assessments (Bailey & Jakicic, 2012).

1. Identify the target or targets to assess.

The first step in developing any assessment is to determine what teachers want to measure. If teachers want to learn about students' learning, it is crucial that the assessment aligns with instruction; this is what makes assessments *valid*. Teacher teams identify the priority standards in a particular unit and create common assessments that specifically measure learning targets derived from those priority standards. Assessment experts and educational consultants Kim Bailey and Chris Jakicic (2012) recommend teachers choose no more than three or four learning targets for each common assessment so they can quickly intervene and correct specific student misconceptions.

2. Determine the level of rigor or depth of knowledge (DOK).

When teachers identify priority standards, they come to consensus on the level of rigor or cognitive demand they expect students to obtain for proficiency. This allows teams to guarantee all students will receive the same level of instruction and consistent teacher expectations across a class, course, or grade level. When developing a common assessment, teams intentionally create items at the DOK levels the teams identify.

When deciding on DOK levels for each target, teams should consider including multiple items at differing levels. If teachers expect proficiency at DOK 3, certainly include items at DOK 3. However, including items on the same target at DOK levels 1 or 2 may help teachers diagnose where student misunderstandings exist.

3. Decide on item types and number of items.

The specific learning target and the team-identified DOK levels for proficiency are instrumental in deciding on an appropriate assessment-item type. For example, selected-response items (multiple choice, true/false, matching) are well suited for DOK levels 1 and 2 (recall and basic skills). However, to accurately assess learning targets at DOK 3 (strategic thinking and reasoning), it is more appropriate to design short-answer or constructed-response items.

There are occasions when multiple choice is the most logical item type to use on a common assessment. In these cases, it is important for teams to keep in mind the power of incorrect responses. Teacher teams can anticipate common mistakes and intentionally create distractor choices that reveal specific student misunderstandings. These items are not intended to trick students, but to uncover misunderstandings so teachers can intervene and correct those mistakes.

In addition, teams must decide on the number of items for each learning target, which leads to test *reliability*. If teams include one selected-response item for a learning target, a student may guess the correct answer. In that case, teachers would not truly know whether the student has mastered the target. On the other hand, if teams include three selected-response items on the same target, they can more confidently assess the student's proficiency level.

4. Consider logistics.

The last step focuses on logistical considerations, with questions like, How long will it take students to complete the test? How will we score the test, What academic vocabulary will we use? How will we clarity the directions? Teams who are intentional about logistics from the beginning find more reliable results in the end.

"A team should learn something every time they write an assessment—learn how other teammates teach the target, learn what proficiency looks like, learn misconceptions students might have, learn how to write good questions."

—Kim Bailey and Chris Jakicic

The ultimate purpose of common assessments is for teachers to learn about their students' learning. Teacher teams can achieve this goal by creating valid and reliable common assessments and using results to adjust instruction and correct student misconceptions. By engaging in these practices, teachers improve their instruction, students deepen their learning, and schools increase the likelihood that students will be successful overall.

References

Bailey, K., & Jakicic, C. (2012). *Common formative assessment: A toolkit for Professional Learning Communities at Work*. Bloomington, IN: Solution Tree Press.

Bailey, K., & Jakicic, C. (2018). *Make it happen: Coaching with the four critical questions of PLCs at Work*. Bloomington, IN: Solution Tree Press.

Reeves, D. (Ed.). (2007). *Ahead of the curve: The power of assessment to transform teaching and learning*. Bloomington IN: Solution Tree Press.