Aligned Assessments

STUDY MODULE 8 **MARCH, 2011**

"The nation is over-tested, but we are under-assessed... Research over the past decade has presented an overwhelming case in favor of providing feedback to students that is frequent, specific, and accurate." (Douglas Reeves)

"On the surface, it would appear that the most important assessments in schools are the annual standardized tests. After all, they command the attention of the president of the United States... But the fact is that these politically important tests pale in their contribution to school success when compared to the assessments teachers develop, administer, and use day to day in the classroom." (Rick Stiggins)

Aligning Instruction and Assessment to Power Standards

If instruction is aligned to assessment, and if instruction is focused primarily on standards, then assessments aligned to the standards provide teachers with credible evidence as to student attainment of those essential standards. And this is precisely what teachers want to know: "How are our students doing relative to the standards that we have determined are essential and will be most heavily assessed on the state exams?" This clear purpose is what motivates extraordinarily busy teachers to take the time to write or design their own assessments that directly match what they are teaching—in particular, those prioritized standards that will enable students to succeed not only on the high-stakes assessments but also in subsequent levels of schooling and in life itself.

Teachers do much of this work in isolation. On their own, they plan standards-based instruction for their students, select or develop appropriate assessments, and then evaluate student progress according to those assessment results. If they conscientiously align their assessments to their instruction, the assessment data can provide insights as to the degree of instructional effectiveness, particularly if they analyze those results and adjust instruction accordingly. In addition, they must analyze student performance data from state and district assessments to determine to what degree students are achieving the standards. Although necessary, all these steps and accountability measures require considerable time, energy, planning, and reflection.

Ultimately, the goal of every teacher is for students to be successful. Carefully designing, adapting, and planning our assessments can lead to greater success for our students.

Collaboration offers many solutions in the face of this very real dilemma. More and more teachers are acknowledging the vital need to "work smarter, not harder." If all the teachers in a particular grade level or department are teaching the same standards in their individualized instructional programs according to their pacing guides or curriculum maps, why not cooperatively plan common formative assessments aligned to those standards? Teachers can design such common assessments to also align with the district quarterly benchmark assessments, end-of-course assessments (secondary), and the state assessments. If each educator brings to the group work his or her own individual insights and experience, every member contributes to and benefits from the collective wisdom shared. The synergistic thinking sure to emerge during such professional collaboration

will produce quality assessments to measure what the teachers want to find out: "How did our students do relative to the standards in focus?"

How often do teachers collaboratively plan a common formative assessment? A recommended interval is once every month, but many find that meeting six or more times during a school year is more realistic.

Because the assessment is *formative,* the participating teachers will usually choose to meet after the pre-assessment is given to all students so as to determine instructional strategies they will each use to prepare students for greater success on the common formative post-assessment.

In addition to aligning the assessments with the state curriculum standards, teachers should also consider Depth of Knowledge (DOK) levels. Collaboration on the design of questions at varying levels of DOK helps to develop stronger assessments. Using blueprints for state tests or sample test questions can help ensure that assessments are aligned with state tests.

Ultimately, the goal of every teacher is for students to be successful. Carefully designing, adapting, and planning our assessments can lead to greater success for our students.

From *Common Formative Assessments: How to Connect Standards-Based Instruction and Assessment* by Larry Ainsworth and Donald Viegut, 2006.

A Quick Refresher About Depth of Knowledge (DOK)

What is DOK?

An alignment model created by Norman Webb, an education research scientist. The purpose is to ensure that there is solid correspondence between the intended learning in the language of the standard and the required demonstration of student learning.

Why is DOK important?

Many states use DOK levels to describe test items and inform interpretation and teaching of the standards. DOK levels and descriptors assist in developing rigorous tasks that promote critical thinking.

Notes about Webb's Model and the different DOK Levels:

- Not the same as difficulty; it is about **intended outcome** *not* difficulty
- Refers to the **complexity** of mental processing that must occur to answer a question, perform a task, or generate a product; it is descriptive, not a taxonomy
- Focuses on how deeply a student has to know the content in order to respond
- Test a **variety** of levels of thinking for complete, **balanced** assessment

What are the Different DOK Levels?

DOK Level 1 (Recall and Reproduction)
Recall or recognition of a fact and/or information (definitions, terms, dates, etc.) required to complete a problem or task.

DOK Level 2 (Basic Application of Skills and Concepts)
Using information, using conceptual knowledge, selecting and/or applying appropriate procedure(s); includes the engagement of some mental processes beyond recalling or reproducing a response; requires a decision as to how to approach a problem or task.

DOK Level 3 (Strategic Thinking)
Requires reasoning, developing a plan for approaching a problem or task, decision making, justification, and complex thinking.

DOK Level 4 (Extended Thinking)
An investigation or application that requires time to research, think, and process multiple conditions of the problem or task; includes non-routine manipulation or connections across disciplines/content areas/multiple sources.

Many standards have the potential of being taught to higher levels. Classroom lessons and assessments should engage students in tasks at the DOK 3 and DOK 4 levels. Requiring students to explain and justify problem solving steps or reasoning on low-level items or questions has the potential of pushing students to demonstrate understanding at DOK levels 2 or 3.

From *Common Assessments Quick Guide* by Washoe County School District, 2009

Guiding PLC Discussion Questions

1. How can you work together to develop assessments that are aligned to standards, state tests, and DOK?
2. Describe assessments you've used or developed that are aligned with standards, state tests, and DOK.
3. What resources do you use for developing good assessments? Are there websites or tools you can share with the group?
4. How can you incorporate sample state test questions or test blueprint information?
5. How can you create and use assessments or assessment items from different DOK levels? Work together to create questions on the same topic that are tied to differing levels of thinking or complexity.
6. How can you adopt or modify a current common assessment to be more aligned with standards, state tests, or DOK? Are there questions you could change (like the example item on page 3) to increase cognitive demand?
7. How do you avoid falling into the trap of "teaching to the test?"

Standards and Tests: Keeping Them Aligned

Recent studies of alignment between tests and standards demonstrate that alignment can be measured reliably by using agreed-upon criteria and procedures. While their methods may differ, generally the studies ask these questions about alignment:

Does the test's content match the content (topics and skills) in the standards? In other words, each test item should correspond to an objective in the standards. Similarly, key ideas in the standards should appear on the tests.

Do the tests and standards cover a comparable "range" or breadth of knowledge, and is there an appropriate "balance" of knowledge across the standards? Alignment studies look at whether a test fairly and effectively samples across the range of objectives described in a state's standards instead of focusing on only a few objectives or disproportionately sampling students' knowledge of some objectives but not others.

Does the level of cognitive demand or challenge called for in the standards match that required for students to do well on the assessment? For example, if the standards require students to synthesize information and explain their thinking, but the test

items only ask students to recall facts, the standards and tests would not be well aligned.

Does the test avoid adding material that is irrelevant to the standard supposedly being assessed? For example, a test item may have an inappropriate "source of challenge," requiring a student to read and understand a long passage about space travel, when it is seeking to measure a student's knowledge of how to estimate distances and travel times.

> Many assessments focus disproportionately on simpler standards. A five state review of English and math standards and tests by Achieve, Inc., concluded "The most challenging standards and objectives are the ones that are under-sampled or omitted entirely . . . [and those] that call for high-level reasoning are often omitted in favor of much simpler cognitive processes."

How Much Alignment Is Enough?

While there is no easy answer to this question, recent studies have yielded a remarkably consistent set of findings about the need to improve standards-to-test alignment. According to researchers, many content standards are too vague to determine whether a particular test item corresponds to the standard or not. Andrew C.

Porter of the University of Wisconsin-Madison, for example, found that states with separate standards for each grade generally had tighter alignment than states whose standards covered a range of grades, such as 3–5, 6–8, and 9–12. In some instances, state standards may not be sufficiently specific to allow an assessment to be aligned tightly with them or to provide adequate guidance for teachers.

But it is possible to go too far in the direction of specificity. Some states list so many standards that tests cannot measure them all. Both Norman L. Webb of the University of Wisconsin-Madison[7] and Achieve, Inc.,[8] a Washington-based nonprofit organization, have found that it is harder for tests to reflect the full range of knowledge included in state standards if content expectations are spread across a large number of standards. Studies also have found that many tests fail to measure adequately the cognitive complexity or "depth of knowledge" described in state standards. For example, in an analysis of 8th-grade math standards and tests in one state, Porter found that the test questions concentrated on length and perimeter, area and volume, and the performance of routine procedures. But the tests generally neglected more complex topics and cognitive demands included in the standards, such as angles or solving non-routine problems.

Small Changes in Test Items Can Shift Cognitive Demand
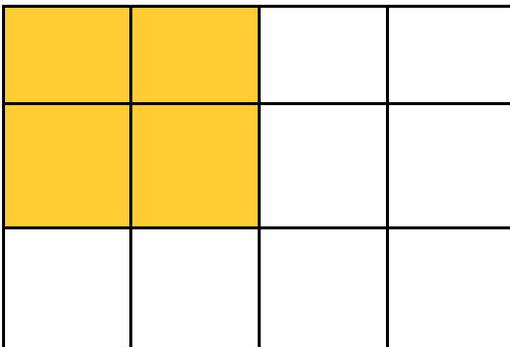
Requires less thinking because it explicitly defines the square units to make the shaded area easier to measure.
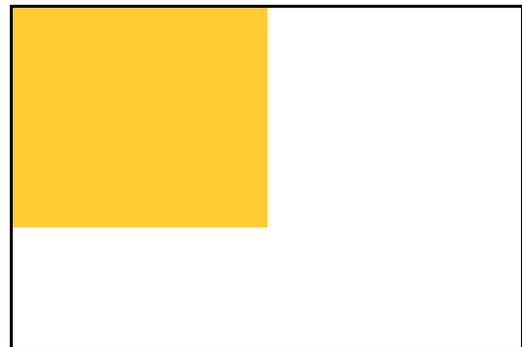
Increases the challenge by requiring that the student make measurements to determine what portion of the figure is shaded.

Putting it into Action: Developing Aligned Assessments

From *RMC Research Corporation, 2000.*

Assessments can draw upon a myriad of testing or measurement tasks ranging from informal, teacher designed activities to standardized, norm referenced tests. The fundamental role of assessment is to provide meaningful feedback for improving student learning, instructional practice, and educational options.

Categories of Assessments

There are many ways to categorize assessments. For our purposes, we will divide the categories into

- selected response assessments,
- constructed response assessments, and
- performance assessments.

Selected response assessments include multiple choice, true/false, and matching tests. They are efficient ways of measuring knowledge acquisition and it is easy to set levels for performance (e.g., 18 correct out of 20 = B). However, you can't always tell whether a student is guessing, his/her depth of knowledge, and/or ability to apply or transfer knowledge. An example of a selected response assessment of Internet-based instruction may be to have students take a written test on facts they were able to access using the web or identifying the right sequence of steps to accomplish something.

Constructed response assessments include fill in the blank, short answers, show your work, and visual depiction activities. Students create answers to questions or prompts. These give teachers a better sense of how well students can convey information and demonstrate some skills like mapping, graphing, and so on. An example of a constructed response assessment of Internet-based instruction may be to ask students to map their journeys in trying to answer a complex question, use data to construct a graph, and/or provide a brief description of the results of a search.

Performance assessments generally require students to demonstrate something that meets specific criteria. This could include, for example, demonstrations of how to conduct a search, posting artwork on the web, producing a book, enactments, scientific demonstrations or exhibits, research papers, and so forth. Some of these are projects, some are performance tasks, and some are culminating activities. Typically, performance assessments illuminate students' skills, conceptual understandings, ability to apply knowledge and skills, performance execution abilities, and process abilities. An example of a performance assessment of Internet-based instruction may be to ask students to post an essay written by accessing information from a number of websites; create a book for younger students on fun websites to visit; and/or use information accessed to conduct science experiments, communicate with students in another country to analyze the implications of international trade agreements, and so forth.

Adapting Assessments

Sometimes you can take an existing assessment and adapt it or customize it for your purposes. When you do this, be sure that you give credit to the source and that you do not inadvertently change the assessment so much that it is ineffective or unreliable. For informal assessments, reliability is not as much of an issue, but if you adapt a scale, you may be changing something that violates its internal validity or reliability check. If you change something, it's a good idea to ask someone else to look at it to see whether it is measuring what you intend to measure.

Developing Assessments

This is not as easy as it may look! Good assessments must:
1. be aligned with specific standards and learning targets;
2. have adequate breadth and depth;
3. affect what is important for students to know and be able to do;
4. be fair and equitable for all students (that is, not reflect cultural, gender, ethnic, or other biases);
5. be aligned with instruction;
6. have appropriate rubrics or scoring criteria attached that already distinguish between levels of performance;
7. be doable within the specified timeframe;
8. be valid and reliable for the purposes for which they are to be used (that is, measure what they say they are measuring and provide consistent results over time and across groups);
9. be readily understood by students; and
10. give information that is useful for student or instructional improvement.

To design an assessment:

Decide on format (selected response, constructed response, performance assessment).

Specify the purpose of the assessment and the standard(s) it will measure.

Specify performance expectations/descriptions and develop rubrics or other indicators for each level of performance.

Clearly articulate the performance task and check that it is aligned with the lesson, measures what is important, is fair, doable within the timeframe, etc.

Questions to Consider

1. Has the assessment accommodated the needs of ALL students, such as offering diverse presentation styles, formats, or timelines to help them achieve high standards?
2. How can the assessment guide instructional strategies for follow-up or future lessons?
3. How can the assessment foster review and improvement in student work?