

DMT INSIGHTS

Issue 22 · 2026 · by Jonathan Brendefur, PhD and Associates



UNDERSTANDING STATE MATHEMATICS ASSESSMENTS: WHAT THEY MEASURE AND HOW TEACHERS CAN USE THEM

Purposes, cognitive demand, assessment categories, and classroom evidence

Introduction

State mathematics assessments provide information about student achievement relative to grade-level standards. Their results can help schools and districts examine patterns in student performance, but they provide a different type of information from the evidence teachers gather during instruction.

Teachers and school leaders need to understand what state assessments measure, how the tasks are designed, and what the results can tell us about student learning. This information is useful when examined alongside students' classroom work, strategies, and explanations. This overview examines the purposes of state assessments, cognitive demand, assessment blueprints, and the relationships among standards, mathematical content, and classroom instruction.

Key Idea • Three questions guide the interpretation of any assessment: What mathematical knowledge does the task assess? What thinking does it require? What can we learn from the student's response?

What Should Educators Know About State Tests?

State assessments are generally summative. They provide information about student achievement at a particular point in time, often near the end of the school year (American Educational Research Association et al., 2014). Formative assessment involves gathering and using evidence during instruction to determine what students understand and what they need to learn next (Black & Wiliam, 1998).

Both serve important purposes. State assessment results can identify broader patterns in achievement, while classroom assessment provides more immediate information about students' strategies, explanations, and understanding (Shepard, 2019). Neither should be interpreted without considering the mathematical content, the assessment tasks, and the context in which the evidence was collected.

State mathematics assessments use a variety of response types. Selected-response items include multiple-choice (single- or multiple-select), drop-down menus, inline choice, and match table grids. Constructed-response items

include equation or text entry, drag-and-drop tasks, fraction models, written explanations scored with rubrics, multipart questions, and graphing tasks.

Selected-response and constructed-response items can assess different aspects of mathematical knowledge, depending on how they are designed. A multiple-choice item may require students to analyze a representation or compare mathematical relationships, while a constructed-response item may assess a routine calculation. The response format alone does not determine the mathematical understanding or cognitive demand required. A task's cognitive demand can also change as it is implemented in the classroom (Stein et al., 2009).

What Can State Assessment Results Tell Us?

State assessments provide information about student achievement across classrooms, schools, and districts. They can help identify performance patterns across mathematical domains and grade levels. However, an overall score does not explain which mathematical relationships a student understands or why a student answered a particular item incorrectly.

National and international assessments provide additional information for different purposes. NAEP examines achievement patterns in the United States, while TIMSS and PISA provide international comparisons using different assessment frameworks and populations (OECD, 2019).

Teachers and school leaders should examine each assessment's purpose, content, and limitations before interpreting the results. State assessment data can help identify areas for further investigation, but classroom tasks and student explanations are needed to examine students' mathematical thinking more closely.

Depth of Knowledge and Cognitive Demand

Webb's Depth of Knowledge (DOK) framework describes the cognitive complexity of tasks and assessments. Webb developed it to examine the alignment between expectations for student learning and the assessments used to measure them (Webb, 1997, 2002). The table gives place-value examples at each level.

DOK level	Description	Mathematics example
Level 1: Recall and Reproduction	Recall a fact, recognize a property, or carry out a familiar procedure.	Identify the value of the 4 in 347.
Level 2: Skills and Concepts	Apply mathematical relationships or make decisions about how to represent or solve a problem.	Represent 347 in two different ways using hundreds, tens, and ones, and explain the equivalence.
Level 3: Strategic Thinking	Analyze relationships, develop a justification, or evaluate competing approaches when the task requires nonroutine reasoning.	Investigate whether regrouping place-value units always preserves the quantity. Develop a justification that applies to different numbers and representations.
Level 4: Extended Thinking	Investigate relationships across a sustained task requiring multiple decisions, analysis, and synthesis.	Collect and analyze measurement data over time, compare representations, and develop and defend conclusions about the data.

DOK describes a task's cognitive demand, not its difficulty or the number of steps required. A model, context, or explanation does not automatically raise the DOK level. The classification depends on the mathematical decisions and reasoning required. Asking students to explain a procedure does not automatically make a task DOK 3. The level depends on the mathematical reasoning and decisions required.

Teachers can include tasks at each level in classroom instruction, including extended investigations that are difficult to include in a timed assessment. DMT Insight 44 examines Depth of Knowledge in more detail.

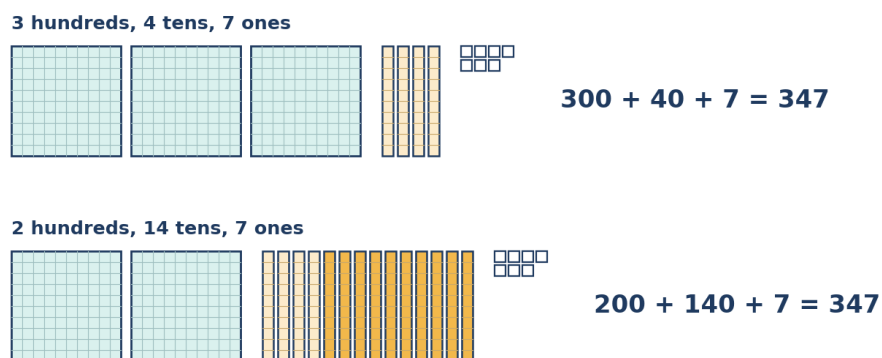
DMTI's Assessment Categories

DMTI uses assessment categories to examine different aspects of students' mathematical knowledge. These include skills and procedures, contextual problem solving, conceptual understanding, and reasoning and justification. These categories help teachers examine students' accuracy with procedures, ability to solve contextual problems, understanding of mathematical relationships, and reasoning. Webb's DOK framework serves a related but different purpose: it describes a task's cognitive demand. A DMTI Level 2 Conceptual task, for example, is not automatically a DOK 2 task. Each framework examines a different feature of the assessment.

The four tasks below use the same third-grade place-value content so that the categories can be compared directly.

DMTI category	Task
Level 1: Skill/Recall	What is the value of the 4 in 347?
Level 2 Problem Solving (L2P)	A school has 347 pencils, packaged in boxes of 100, packs of 10, and single pencils. If one box of 100 pencils is opened and repackaged into packs of 10, how many boxes, packs, and single pencils are there now? Explain how you know the total has not changed.
Level 2 Conceptual (L2C)	Represent 347 using hundreds, tens, and ones. Show another way to compose 347 by regrouping one hundred into ten tens. Explain why the total remains the same.
Level 3: Reasoning and Justification	One student says, "If I trade one hundred for ten tens, I have more because I have more pieces." Another says, "The number of pieces changes, but the quantity stays the same." Evaluate their reasoning and justify your conclusion using place-value relationships.

The Level 2 Conceptual task asks students to decompose one hundred into ten tens and explain why 3 hundreds, 4 tens, and 7 ones and 2 hundreds, 14 tens, and 7 ones represent the same quantity (Figure 1).



One hundred is regrouped as ten tens (darker). The number of pieces changes; the quantity stays 347.

Figure 1. Decomposing one hundred into ten tens changes the number of units being counted, but the total quantity remains 347.

Key Idea • DMTI assessment categories describe different aspects of mathematical knowledge. Webb’s DOK describes a task’s cognitive demand. Both are useful when examining assessment items.

Assessment Blueprints and Instruction

State assessment blueprints describe how an assessment samples grade-level standards and mathematical content. Depending on the assessment, they may identify reporting categories, item types, scoring, and expectations for mathematical reasoning.

Instruction should be guided by grade-level standards, mathematical learning progressions, and the relationships students need to understand. Assessment blueprints provide additional information about how those standards are sampled on a particular state test. They should inform instructional planning without narrowing instruction to particular item formats.

Blueprint specifications differ across assessment programs and change over time. Teachers and school leaders should consult the current blueprint for their state’s assessment and identify the program and year when discussing specific features.

Connecting Mathematical Content Across Domains

Mathematics assessment items may require students to use relationships from more than one domain. For example, determining the area of a rectangle involves multiplication and geometric measurement. Interpreting a line plot of fractional measurements requires students to understand number lines, units, and data.

These connections should be part of instruction because mathematical ideas are related, not simply because they may appear together on a state assessment.

Summary

State mathematics assessments provide information about achievement relative to grade-level standards. They can help teachers and school leaders identify patterns in student performance, but they do not fully explain students' mathematical understanding.

Classroom tasks, student strategies, and explanations provide additional evidence. By examining this evidence alongside state results, teachers can identify mathematical relationships that need further development and plan instruction that supports both conceptual understanding and procedural fluency.

References

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. American Educational Research Association.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7–74. <https://doi.org/10.1080/0969595980050102>
- Organization for Economic Co-operation and Development. (2019). *PISA 2018 results (Volume I): What students know and can do*. OECD Publishing.
- Shepard, L. A. (2019). Classroom assessment to support teaching and learning. *The ANNALS of the American Academy of Political and Social Science*, 683(1), 183–200. <https://doi.org/10.1177/0002716219843818>
- Stein, M. K., Smith, M. S., Henningsen, M. A., & Silver, E. A. (2009). *Implementing standards-based mathematics instruction: A casebook for professional development* (2nd ed.). Teachers College Press.
- Webb, N. L. (1997). *Criteria for alignment of expectations and assessments in mathematics and science education* (Research Monograph No. 6). Council of Chief State School Officers.
- Webb, N. L. (2002). *Depth-of-knowledge levels for four content areas [Unpublished manuscript]*—Wisconsin Center for Education Research.

Social Media

What do state mathematics assessments tell us about student learning?

State assessments provide information about achievement relative to grade-level standards. Classroom assessments allow teachers to examine students' strategies, representations, explanations, and understanding more closely. This DMT Insight examines:

- What state assessment results can and cannot tell us
- How Webb's Depth of Knowledge describes cognitive demand
- How DMTI's assessment categories differ from DOK levels
- How teachers can use blueprints alongside standards and classroom evidence

Read the full Insight at mathsuccess.io/dmtinsights.

What evidence do you use, alongside state results, to plan instruction?

