



Developing Mathematical Thinking:
Standardized Testing: Insights and Tools

RESEARCH OVERVIEW

BY

Jonathan Brendefur, PhD
and Associates

STANDARDIZED TESTING: INSIGHTS AND TOOLS FOR EDUCATORS

(RESEARCH OVERVIEW)

Introduction

Standardized assessments are critical educational tools, yet educators and administrators often misunderstand or underutilize them. This research overview addresses five key questions about summative state tests, providing detailed insights and recommendations to improve instructional alignment and student outcomes. Additionally, it incorporates supporting information to clarify misconceptions, explore assessment frameworks, and emphasize content integration across domains.

What Should Educators Know About State Tests?

State tests are summative assessments that differ significantly from formative assessments in their purpose and design. **Summative assessments** evaluate student learning at the end of an instructional period, typically after the school year, and are often high-stakes measures used to compare performance against benchmarks or standards. In contrast, **formative assessments** are embedded within daily instruction and serve as low-stakes tools to monitor student progress, identify misconceptions, and provide actionable feedback for both teachers and students. Formative assessments focus on improving learning during the instructional process, while summative assessments aim to measure the outcomes of that learning (AERA, APA, & NCME, 2014; Dixson & Worrell, 2016).

State-level mathematics assessments utilize a variety of response types to measure student understanding. **Selected-response items** include multiple-choice (single- or multiple-select), drop-down menus, inline choice, and match table grids. **Constructed-response items** require equation editor/text entry, drag-and-drop tasks, fraction models, written explanations scored with rubrics, multipart questions combining multiple formats, and graphing tasks. Both classroom-based assessments and state tests include tasks that assess procedural and conceptual understanding, though they often differ in emphasis and format. Classroom assessments frequently incorporate open-ended tasks, such as modeling and reasoning activities, that promote deeper conceptual understanding and allow students to demonstrate their thinking processes. On the other hand, state tests often rely on selected-response items that emphasize procedural skills due to their efficiency in scoring and standardization. Research suggests that when students engage in cognitively demanding tasks during classroom instruction—such as reasoning through complex problems, applying knowledge in new contexts, and justifying their solutions—they develop foundational skills that not only improve their understanding but also transfer effectively to both formative classroom assessments and summative state tests (Stein et al., 2009; Shepard, 2019).

What Misconceptions Exist About Standardized Testing?

Misconceptions about standardized tests often stem from frustration with their perceived irrelevance or overemphasis on rote skills. Many educators view these tests as disconnected from meaningful learning, focusing excessively on accountability rather than instructional improvement. However, standardized tests serve broader purposes, such as ***evaluating systemic performance and equity across schools, districts, and states***.

National and international assessments serve distinct purposes that provide valuable insights into student achievement. For instance, national assessments, such as the National Assessment of Educational Progress (NAEP), offer data on academic trends across various demographic groups, helping to identify disparities and inform policy decisions. International assessments, such as the Programme for International Student Assessment (PISA) and the Trends in International Mathematics and Science Study (TIMSS), benchmark national education systems against global standards by evaluating students' real-world problem-solving skills.

To address common misconceptions about standardized testing, educators should recognize that these assessments complement formative assessments by offering a broader picture of student achievement beyond the classroom. Professional development opportunities can further support teachers by helping them understand how formative instructional practices build the higher-order thinking skills necessary for success on state and international exams.

How Do Depth of Knowledge (DOK) Levels Inform Assessment Design?

Norman Webb's Depth of Knowledge (DOK) framework, developed by Webb (1997, 2002), categorizes tasks into four levels of cognitive complexity to evaluate the rigor of learning objectives and assessments. The first three levels—**Recall, Skills/Concepts**, and **Strategic Thinking**—are commonly integrated into state standardized tests. In contrast, the fourth level, **Extended Thinking**, is typically reserved for classroom-based projects requiring sustained analysis over time.

1. **Recall (DOK Level 1):** This level involves basic retrieval of facts or procedures. An example could be identifying the sum of $7+5$ or naming a shape, such as a triangle or rectangle. These tasks require straightforward recall without deeper reasoning.
2. **Skills/Concepts (DOK Level 2):** At this level, students engage in tasks requiring some mental processing beyond a rote response. For instance, they might use a number line to find the difference between two numbers, such as solving $15-8$ by counting backward. Another example could involve comparing the lengths of two objects using non-standard units (e.g., paperclips) or interpreting a simple bar graph to answer questions like, "How many more apples than oranges are there?"
3. **Strategic Thinking (DOK Level 3):** This level demands analysis, synthesis, or justification. An example could involve explaining why two fractions are equivalent using visual models, such as demonstrating using a number line or bar model that $\frac{1}{2}$ is the same as $\frac{2}{4}$.

State tests predominantly assess DOK Levels 1 and 2, constituting approximately 80–90% of items, with only 10–20% targeting Level 3. Extended Thinking (DOK Level 4), which involves complex, real-world tasks like designing a school event budget using percentages, is rarely included in state assessments due to time constraints and logistical challenges (Herman et al., 2012). For example, a RAND Corporation analysis of over 5,100 state test items found that fewer than 5% of mathematics items reached DOK Level 4, with most assessments prioritizing procedural fluency over extended reasoning (Herman et al., 2012). Several states have begun to require students to complete a

mathematics performance task that often has elements that reach a Level 4 of DOK. However, this process is relatively new to large-scale assessment, and data are still being collected.

Webb designed the DOK framework to align curriculum, instruction, and assessments, emphasizing that higher levels of cognitive demand, such as Level 4, require students to synthesize information, evaluate solutions, and apply knowledge across contexts over extended periods (Webb, 2002). While most state tests focus on lower DOK levels for standardization, educators can scaffold classroom instruction to include Level 4 tasks, such as analyzing long-term environmental data or designing community projects, to foster deeper critical thinking and real-world problem-solving skills.

What is the Relationship Between Assessment Blueprints and Instructional Priorities?

Assessment blueprints play a critical role in aligning instructional priorities with the structure and expectations of state tests. These blueprints detail the components of standardized assessments, including item types, cognitive demands, and the integration of multiple standards within tasks. By providing a clear framework, blueprints help educators ensure that their curriculum and instruction prepare students for the specific demands of state assessments. For example, the Smarter Balanced Assessment Consortium (SBAC) and Partnership for Assessment of Readiness for College and Careers (PARCC) blueprints frequently combine multiple standards within a single task, requiring students to apply knowledge across domains rather than focus on isolated skills. This integrated approach reflects the complexity of real-world problem-solving and encourages teachers to design lessons that mirror these expectations.

Based on blueprint specifications, teachers could design tasks at varying levels from the DMTI assessment framework in third-grade place value instruction. At Level 1 (Skill/Recall), students might identify the tens digit in a number like 243. At Level 2, tasks can be differentiated: for Level 2 Problem Solving (L2P), students could solve contextual problems involving combinations of place value units, such as determining multiple ways to compose a given number using tens and ones (e.g., finding three different combinations of boxes, packs, and singles for 318 pencils). For Level 2 Conceptual (L2C), students might use base-ten blocks or visual models to compose a number like 345 in several ways, demonstrating their understanding of the relationships between units. At Level 3 (Reasoning and Justification), students would be asked to communicate and critique reasoning—for example, analyzing two different students' models for 247 and justifying which is closer to 200 or 300 using a number line, or explaining why one representation is more efficient or accurate than another. This structure allows teachers to assess procedural skills, problem-solving, conceptual understanding, and students' ability to reason and justify their thinking, all of which are emphasized in the DMTI assessment work.

Using assessment blueprints involves more than simply aligning content; it also requires careful attention to cognitive demands at different DOK levels. Professional development workshops can support teachers in analyzing blueprints and designing high-quality assessment items that reflect state standards. These workshops often include professional development on interpreting blueprint specifications, identifying key standards, and creating tasks that balance procedural fluency with conceptual understanding. By leveraging these tools, educators can ensure that their instruction meets state testing requirements and promotes meaningful student learning experiences.

How Do Assessments Integrate Content Across Domains?

State tests frequently assess integrated content rather than isolated standards, emphasizing the interconnectedness of mathematical concepts and preparing students for complex problem-solving scenarios. For example, a PARCC task

might combine geometry concepts with algebraic reasoning, requiring students to apply their understanding of shapes and spatial relationships while solving equations or inequalities. Similarly, an SBAC item could challenge students to use proportional reasoning with data analysis, such as interpreting a graph to identify trends and make predictions based on ratios or percentages. These integrated tasks encourage students to synthesize knowledge across multiple domains, promoting real-world applications and higher-order thinking skills.

Specific examples illustrate how integration works in practice. A Grade 3 task might ask students to calculate a garden's area using multiplication while visually representing the solution with an area model. This type of task combines foundational arithmetic skills with spatial reasoning, helping students connect abstract calculations to tangible representations. In Grade 4, a task could involve analyzing a bar graph showing apple harvests to determine equivalent fractions and justify reasoning using visual models. Here, students integrate data interpretation with fraction concepts, requiring them to move beyond procedural fluency to explain their thinking meaningfully. Similarly, middle school tasks might involve calculating unit rates for a construction project while graphing results to visualize trends, combining proportional reasoning with statistical analysis.

Integrating content within state assessments underscores the importance of designing classroom lessons that reflect these multidimensional expectations. Teachers can create activities that combine multiple standards, such as using geometric principles to solve real-world measurement problems or analyzing statistical data while applying proportional reasoning. Professional development opportunities can further support educators by providing training on analyzing test blueprints, identifying opportunities for integration, and exploring sample items from PARCC and SBAC that demonstrate the assessment of multiple standards within a single task. By incorporating integrated tasks into daily instruction, teachers can help students develop both procedural fluency and conceptual understanding, ensuring they are prepared for success on state assessments and in real-world problem-solving scenarios.

Conclusion

A balanced assessment system is essential for supporting both immediate learning goals and long-term student success. Classroom-based assessments emphasize deeper learning through open-ended tasks, such as modeling and solving complex problems, which align with higher-order Depth of Knowledge (DOK) levels (3 and 4). In contrast, state-level tests primarily focus on selected-response items that assess procedural fluency at lower DOK levels (1 and 2). Educators should use Webb's DOK framework to scaffold cognitive rigor, emphasize real-world contextual learning, and address misconceptions through error analysis protocols to bridge the gap between classroom practices and standardized test design. Administrators play a crucial role by providing professional development on interpreting test blueprints, advocating for through-year testing models to reduce high-stakes pressure, and investing in resources for underprivileged schools to ensure equitable access to preparatory materials. Educators and administrators can leverage assessments as tools for meaningful educational improvement rather than barriers to innovation by fostering alignment between classroom instruction and state assessments while supporting systemic reforms.

References

American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME]. (2014). *Standards for educational and psychological testing*. Washington, DC: AERA.

Dixon, D. D., & Worrell, F. C. (2016). Formative assessment: A meta-analysis and a call for research. *Educational*

Measurement: Issues and Practice, 35(4), 29–37.

Herman, J., et al. (2012). *Estimating the percentage of students tested on cognitively demanding items through state achievement tests*. RAND Corporation.

Shepard, L. A. (2019). The role of classroom assessment in teaching and learning. In G. T. L. Brown & L. R. Harris (Eds.), *Handbook of human and social conditions in assessment* (pp. 62–78). Routledge.

Smarter Balanced Assessment Consortium [SBAC]. (n.d.). *Mathematics summative assessment blueprint*. Retrieved from Smarter Balanced website.

Stein, M., Smith, M., Henningsen, M., & Silver, E. (2009). *Implementing standards-based mathematics instruction: A casebook for professional development*. Teachers College Press.

Webb, N. L. (1997). *Research monograph No. 8: Criteria for alignment of expectations and assessments in mathematics and science education*. Washington, DC: Council of Chief State School Officers.

Webb, N. L. (2002). Depth-of-knowledge levels for four content areas. *Council of Chief State School Officers*.

WestEd. (2017). *Evaluation of the alignment between the Common Core State Standards and assessments*. WestEd Research Institute.

Organisation for Economic Co-operation and Development [OECD]. (2019). *PISA 2018 results*. Paris: OECD Publishing.

Social Media

Curious about how to align your teaching with state testing requirements, while still supporting deep learning? Dive into the full research overview and discover practical tools and strategies to change your assessment and instruction approach!

How do state tests fit into the bigger picture of student learning? The article "Standardized Testing: Insights and Tools for Educators" by Brendefur et al. offers a clear, research-based overview of what educators should know about summative state assessments—and how to use this knowledge to improve instruction. This overview tackles common misconceptions, explains the difference between summative and formative assessments, and highlights how frameworks like Depth of Knowledge (DOK) and assessment blueprints can guide teachers in designing meaningful tasks.

Why does this matter? Understanding the structure and purpose of standardized tests helps teachers move beyond rote practice and focus on building students' conceptual understanding and problem-solving skills. The article also shows how integrating content across domains and aligning classroom activities with assessment blueprints prepares students for real-world problem solving and success on state and national assessments. Professional development and collaboration empower educators to create lessons that balance procedural fluency with higher-order thinking.



For more information

✉ contact@dmtinstitute.com

📘 [dmtinstitute](#)

🌐 www.mathsuccess.io

🌐 [dmt-institute](#)

📍 Developing Mathematical Thinking Institute, Inc.
500 South 8th Street
Boise, Idaho 83702

📷 [dmtinstitute](#)

🐦 [dmtinstitute](#)

