Instructionally Relevant Assessment Systems

What Is the Role of Performance Assessments?

Aneesha Badrinarayan

LEARNING POLICY INSTITUTE

JUNE 2025

Acknowledgments

The author thanks Sara Cooper, Nathan Dadey, Susan Lyons, Scott Marion, Jim Pellegrino, and Katie Van Horne for their thought partnership through the development of this report, as well as Learning Policy Institute (LPI) colleagues Linda Darling-Hammond and Julie Woods for their guidance, feedback, and encouragement. In addition, the author thanks the members of the LPI Communications team for their invaluable support in editing, designing, and disseminating this report.

This research was supported by the Carnegie Corporation of New York, Chan Zuckerberg Initiative, William and Flora Hewlett Foundation, and Walton Family Foundation. The Heising-Simons Foundation, Raikes Foundation, Sandler Foundation, Skyline Foundation, and MacKenzie Scott provided additional core operating support for LPI. The ideas voiced here are those of the author and not those of funders.

External Reviewers

This report benefited from the insights and expertise of two external reviewers: Raymond Pecheone, Professor of Practice, Emeritus, Stanford University, and Rich Shavelson, Professor and Dean Emeritus, Stanford University. The author thanks them for the care and attention they gave the report.

Suggested citation: Badrinarayan, A. (2025). *Instructionally relevant assessment systems: What is the role of performance assessments?* Learning Policy Institute. https://doi.org/10.54300/411.469

This report can be found online at https://learningpolicyinstitute.org/product/instructionally-relevant-assessment-systems.

Document last revised June 20, 2025

Table of Contents

Executive Summary iii

Introduction 1

Designing Assessments With an Emphasis on Positive Instructional Impact 2

Performance Assessments as a Path to Center Assessment as Learning in Large-Scale Systems 5

Designing High-Quality Performance Assessments for Systemic Use 8

Common Tasks and Rubrics 8

Rubrics and Calibrated Scoring 9

Rigorous Assessment Development Processes 10

Tuning Performance Assessments Based on Purpose 10

Performance Tasks in Action in Large-Scale Programs 14

Recommendations for System Leaders 19

Conclusion 21

Appendix A: Performance Assessment System Details and Sample Tasks 22

Appendix B: Frequently Asked Questions 41

Endnotes 46

About the Author 48

List of Figures and Tables

Figure 1 Assessment Continuum 7

Figure 2 Constants or "Non-Negotiables" for Performance Assessments 11

Figure 3 Knobs to Adjust for Performance Assessments 11

Table 1 Trade-Offs for Assessment Design 4

Executive Summary

Since the rise of state assessments whose primary function is to yield scores that can be used to compare schools and groups of students, most states have developed their state assessment programs under the assumption that either: (a) state tests are not intended to meaningfully shape instruction, or (b), if they are, the information provided in score reports is sufficient to support instruction. Indeed, the prevailing guidance about large-scale assessments is that they should serve a program monitoring role and *not* be used to guide instruction. This approach reflects sound reasoning. It is hard for an external, efficient, infrequent assessment to play a meaningful role in guiding instruction, and many measurement experts suggest that state assessments should be supplemented by other supports, closer to the classroom, to provide real instructional support. State tests have been designed accordingly, making trade-offs that value efficiency and reliability over impact on teaching and learning.

While these recommendations to keep the summative assessment separate from instruction might reflect some conceptions of *best* practice, they unfortunately do not reflect *real* practice. When policymakers and researchers listen to teachers and local leaders, they routinely hear not only that teachers are changing what they teach to better match test content, but that teachers are often encouraged to change instruction in ways that actively trade off features of high-quality learning experiences for those that reflect testing experiences.

While state officials may not intend for their assessments to have this impact, the footprint of state assessments grows each time a district chooses an interim assessment that promises to predict performance on the state test; purchases access to formative assessment item banks with questions that mirror (or reproduce) released state assessment questions; or provides guidance about curriculum choices, scope and sequence, and time for content-specific teaching and learning that is driven by state assessment design and scores. The result is a cascade of signals that position state assessments as a major driver shaping the learning experiences that students have in the classroom.

Faced with this reality, many state leaders are reconsidering their approach to assessment programs. Instead of designing assessment systems under the assumption that state assessments are not influencing teaching and learning—or that those influences are an unfortunate "cost of doing business"—state leaders are asking: If we know teachers and local leaders take cues from state-provided assessments, how can we create instructionally relevant assessments that incentivize shifts toward better teaching and learning?

Designing Assessments With an Emphasis on Positive Instructional Impact

Designing state assessments with instructional impact in mind requires reconsidering what features and values to prioritize in assessment design. For example, many developers of current state assessments view assessment tasks that require human scoring (tasks like constructed response items and student-written essays) as an unnecessary burden. Such tasks cost more in both money and time than single-select multiple choice questions and can require a great deal of coordination and capacity from state education agencies. In addition, the points given for these kinds of items generally get combined with points from much more efficient multiple choice questions in ways that limit the impact of student

performance of human-scored items on test scores or achievement levels. It makes sense then to devalue these kinds of items *if* the only important outcome of the test is a reliable numerical score. If, however, an important outcome of the assessment is building teacher capacity around understanding standards and disciplinary pedagogy (as they are when instructional impact is centered), human-scored performance-based assessment items and tasks become much more valuable. Participating in scoring activities can:

- allow teachers to participate in facilitated professional learning connecting standards to expectations on the assessment;
- provide opportunities to practice analyzing student work;
- encourage teachers to collaborate with colleagues across classrooms, schools, and districts;
- provide examples of the kinds of performances students may need to practice during the course of instruction;
- foster teachers' understanding of their own students' strengths (if scoring their own students' work); and
- disrupt deficit narratives about what their students can and cannot do.

This kind of information is much more valuable for teaching and learning than decontextualized test scores that often prompt teachers to turn to ineffective reteaching and remediation strategies. High-quality performance assessments increase the validity of scores resulting from an assessment by providing better insight into what students know and can do relative to the standards being measured. Moreover, they are a particularly compelling component of assessment programs when leaders shift from centering reliable and comparable scores as the only important outcome of an assessment and shift toward including impact on instruction as an equally important outcome.

When state leaders center instructional impact as just as important to surfacing data that can serve program monitoring functions, performance assessments consistently emerge as an essential element of many large-scale system designs. Done well, performance assessments surface evidence of what students know and can do in deeply authentic and meaningful ways. This leads to better alignment between assessments and state standards like Common Core and the Next Generation Science Standards, and, as a consequence, more valid assessment scores. High-quality performance assessments are also more relevant and meaningful to students than a diet of decontextualized selected-response questions. Relevance can increase student engagement and perseverance through complex tasks and improve the assessments' ability to surface the range of sophisticated understanding that diverse learners may possess. Perhaps most importantly to many state and district leaders, performance assessment can position large-scale assessments as tools that support high-quality teaching and learning by signaling features of effective learning and assessment environments and by providing actual classroom experiences (in the case of curriculum-embedded tasks).

When they are designed appropriately and used in conjunction with other measures, performance assessments can be reliably scored, generating trustworthy and comparable scores at the student and aggregate levels. In this report, the focus is on the use of authentic performance tasks used together with more standardized kinds of assessment, as this is what most systems are exploring. Many large-scale systems—states and districts as well as national and international curricular and assessment programs like Advanced Placement and International Baccalaureate—use performance tasks as part

of their system because they elicit evidence of student thinking that is not readily surfaced through selected-response items and influence instruction in positive ways. When designing performance tasks for use in larger-scale systems, developers and leaders often emphasize certain features of assessment design, implementation, and scoring such as common tasks and rubrics, calibrated scoring, and rigorous assessment development processes that produce tasks that contribute to trustworthy student scores that can be aggregated and compared as needed. These features distinguish performance tasks that can be used to generate trustworthy and comparable scores from those that are often developed locally as part of meaningful instruction and lack features that would allow them to be used within large-scale systems (e.g., a project or task developed by an individual teacher to be used as part of coursework).

Recommendations for System Leaders

Performance assessments can transform assessment systems into forces for improved teaching and learning. Doing so requires that system leaders position performance assessments—and the supports needed for their design and use—as a valued element of both instruction and student performance. As leaders consider how to reorient their assessment systems toward instructional relevance, it may be useful to consider the following recommendations:

- **Demand assessments that measure what matters.** Ensure that assessments actually measure the higher-order thinking and problem-solving, disciplinary practices, and other deeper learning competencies that students need to be ready for college, careers, and citizenship.
- **Recognize the transformative potential of signaling as it shapes student learning experiences.** Large-scale assessment systems frequently make their biggest mark on instruction through their signaling function, influencing decisions about what gets taught, how students experience learning, and what success should look like. Including performance assessments in assessment systems can be transformative doing so encourages instructional shifts toward deeper learning.
- **Leverage performance assessments strategically.** Many large-scale systems that leverage performance assessments do so in conjunction with other assessment instruments, such as on-demand selected-response items. The combination of both approaches to assessment allows assessment designs that can sample wider coverage while still providing substantial and sufficient evidence of students' ability to reason and demonstrate learning in sophisticated ways within and (if appropriate) across disciplines. The key is striking a strategic and effective balance and ensuring that performance assessments count for enough of students' final scores that stakeholders pay attention to the knowledge, skills, and abilities needed to complete these tasks.
- **Make the assessment worth the investment.** A major element of the value proposition of performance assessments lies in their authenticity and educative nature—that is, how assessments build educators' understanding of content and pedagogy in their disciplines. Leaders should prioritize the development and use of authentic, relevant, and sophisticated tasks that motivate students and provide a beacon of what their routine experiences in the classroom should look like. Leaders also should engage all classroom teachers in the development and interpretation of these tasks so that teachers can have access to the rich information about student thinking that such tasks produce and to build support for the pedagogy that enables deep learning.

- **Consider creative resource allocation to support professional learning.** Many systems that center performance assessments emphasize the impact on teaching and learning for students. Leaders should also consider the potential for meaningful and sustained professional learning when making resource allocation and budgeting decisions. While performance assessments cost more to design and score than multiple choice questions, resources can be reallocated from test preparation and interim assessments that are essentially practice tests. Professional development time can also be allocated for design and scoring, as teachers consistently note the benefits of designing and reviewing performance assessments for their own learning and planning. Resources spent on meaningful performance task development and scoring can contribute to improved teacher practice *and* student learning experiences, which is likely to lead to better student outcomes than simply practicing the questions on a superficial final exam.

As leaders in a growing number of states and systems consider how to break the cycle of assessments being used to limit learning opportunities for students, they are considering a different purpose—and a different set of trade-offs—for their assessment system designs. Many leaders are drawing a line in the sand. Teaching and learning are paramount, and any assessment system that does not have a positive impact on teaching and learning cannot be acceptable. When leaders make positive instructional impact a necessary condition of high-quality assessment systems, performance assessments routinely emerge as an important element of system designs. Done well, and often in conjunction with other measures, performance assessments can provide better evidence of what students know and can do while helping students and teachers alike better understand how meaningful instruction should look and feel.

Introduction

Complaints from educators about the negative effects of testing on instruction have been prominent for at least 2 decades, since high-stakes test-based accountability was introduced in federal law. As just one example, during a recent listening tour across multiple states, teachers and district instructional leaders were asked what kinds of changes they make to instruction and why. Although educators were not prompted to consider the role of state assessments, they overwhelmingly shared experiences that highlighted how assessment examples and results that they receive from large-scale external assessments, such as their statewide summative assessments and interim assessments adopted by the district, influence their teaching. For example, teachers reported the following:

- "I cut out the projects and extended investigations—I know my kids will probably enjoy 'being a scientist,' but if I do them, I can't cover enough of the content in the standards for them to get the right answers on the test."
- "My curriculum has a lot of students talking to each other—Socratic seminars, social work time, things like that. They're never going to be able to talk to each other on a test, so I usually cut down on those. They need to practice the kinds of stuff they're asked on the tests."
- "On the test, kids have to drag and drop text boxes, so I change our classroom assessments that ask for drawing or adding to models to match what they have to do on the state test."
- "I know we talk about practices and application, but people judge not only me but my kids based on whether they can do what's on the state test, which is about the facts. I add test review, flashcards, *Jeopardy!*—those kinds of things. There isn't enough time, but it's important that my kids feel prepared for the test."

These changes to instruction are sobering. Teachers frequently share that they change high-quality instructional materials and approaches to better match the tests. These changes typically serve to limit opportunities for deep, engaging, and meaningful learning.

Since the rise of state assessments whose primary function is to yield scores that can be used to compare schools and groups of students, most states have developed their assessment programs under the assumption that either: (a) state tests are not meaningfully shaping instruction, or (b), if they are, the information provided in score reports is sufficient to support instruction. Indeed, the prevailing guidance about large-scale assessments is that they should serve a program monitoring role and *not* be used to guide instruction.[1] This approach reflects sound reasoning: It is hard for an external, efficient, infrequent assessment to play a meaningful role in guiding instruction, and many measurement experts suggest that state assessments should be supplemented by other supports, closer to the classroom, to provide real instructional support. State tests have been designed accordingly, making trade-offs that value efficiency and reliability over impact on teaching and learning.

While these recommendations to keep summative assessments separate from instruction might reflect *best* practice, they unfortunately do not reflect *real* practice. When states listen to teachers and local leaders, they routinely hear not only that teachers change what they teach to better match test content, but that teachers are often encouraged to change instruction in ways that actively trade off features of high-quality learning experiences for those that better match efficient testing experiences. While states

may not intend their assessments to have this impact, the footprint of state assessments grows each time a district chooses an interim assessment that promises to predict performance on the state test; purchases access to formative assessment item banks with questions that mirror (or often reproduce) released state assessment questions; or provides guidance about curriculum choices, scope and sequence, and time for content-specific teaching and learning that is driven by state assessment design and scores. The net result is a cascade of signals that position state assessments as a major driver for the kinds of learning experiences that students have in the classroom (see Table 1).

Faced with this reality, many states are reconsidering their approach to assessment programs. Instead of designing assessment systems under the (false, and known to be so) assumption that state assessments are not influencing teaching and learning—or that those influences are an unfortunate "cost of doing business"—state leaders are asking: If we know teachers and local leaders take cues from state-provided assessments, how can we use that power to create assessments that incentivize shifts toward better teaching and learning?[2]

State leaders are asking: If we know teachers and local leaders take cues from state-provided assessments, how can we use that power to create assessments that incentivize shifts toward better teaching and learning?

Designing Assessments With an Emphasis on Positive Instructional Impact

Designing state assessments with instructional impact in mind requires reconsidering what features and values to prioritize in assessment design. For example, many current state assessments view tasks that require human scoring (tasks like constructed response items and student-written essays) as an unnecessary burden. They cost more in terms of both money and time to score than single-select multiple choice questions and can require a great deal of coordination and capacity from state education agencies. In addition, the points given for these kinds of items generally get combined with points from much more efficient multiple choice questions in ways that limit the impact of student performance of human-scored items on test scores or achievement levels. It makes sense then to devalue these kinds of items *if* the only important outcome of the test is a reliable score. If, however, an important outcome of the assessment is building teacher capacity around instructional impact by helping them understand the standards and their students' responses, human-scored assessment items become much more valuable. Participating in scoring activities can:

- allow teachers to participate in facilitated professional learning connecting standards to expectations on the assessment;
- provide opportunities to practice analyzing student work;
- encourage teachers to collaborate with colleagues across classrooms, schools, and districts;
- provide examples of the kinds of performances students may need to practice during the course of instruction;
- foster teachers' understanding of their own students' strengths (if scoring their own students' work); and
- disrupt deficit narratives about what their students can and cannot do.

This kind of information is much more valuable for teaching and learning than decontextualized test scores that often prompt teachers to turn to ineffective reteaching and remediation strategies. High-quality performance assessments increase the validity of scores resulting from an assessment by providing better insight into what students know and can do relative to the standards being measured. Moreover, they are a particularly compelling component of assessment programs when leaders shift from centering reliable and comparable scores as the only important outcome of an assessment and shift toward including impact on instruction as an equally important outcome. Table 1 describes the ways different perspectives on assessment goals manifest in different assessment designs, illustrating why certain approaches may be more useful for different goals and priorities. Figure 1 shows a continuum of assessments from traditional tests to assessments of deeper learning.

When state education departments view instructional impact as just as important as producing data to serve program monitoring, performance assessments consistently emerge as an essential element of large-scale system designs. Done well, performance assessments surface evidence of what students know and can do in deeply authentic and meaningful ways. This leads to better alignment between assessments and state standards like Common Core and the Next Generation Science Standards, and, as a consequence, to more valid assessment scores. High-quality performance assessments are also more relevant and meaningful to students than a diet of decontextualized selected-response questions. Relevance can increase student engagement and perseverance through complex tasks and improve the assessments' ability to surface the range of sophisticated understanding that diverse learners may possess. Perhaps most importantly to many state and district leaders, performance assessment can position large-scale assessments as tools that support high-quality teaching and learning by signaling features of effective learning and assessment environments and by providing actual classroom experiences (in the case of curriculum-embedded tasks). When they are designed appropriately and used in conjunction with other measures, performance assessments can be reliably scored, generating trustworthy and comparable scores at the student and aggregate levels.[3]

This report defines performance assessments and discusses what features of performance assessments make them particularly compelling to system leaders who are centering instructional impact in their assessment system design. It also discusses major features of performance assessments that are intended to be used as part of assessment designs that generate trustworthy and comparable student scores. Finally, it highlights existing systems that leverage performance assessments as part of high-stakes assessment designs.

Table 1. Trade-Offs for Assessment Design

Feature	**Perspective A:** The primary focus of state assessment design should be on program monitoring. Positive instructional impact is nice to have but not a must-have.	**Perspective B:** The primary focus of state assessment design should be on positive instructional impact. The assessment should also surface program monitoring information, but positive impact on instruction is a non-negotiable.
Information prioritized	• Answering the question: Are students proficient in the standards? • Coverage of as many standards as possible • Aggregate scores that can be used for comparing overall performance for schools, districts, and student subgroups at the macro level • Achievement/performance-level descriptors relative to grade-level standards	• Answering the question: How well do students understand and use the standards? • Sufficient but not comprehensive coverage, emphasis on opportunities to measure depth, disciplinary practices, and deeper learning • Individual and aggregate scores as well as examples of tasks and student work that help educators and leaders make shifts that improve instruction • Achievement/performance-level descriptors that more precisely pinpoint what students know and can do (e.g., along multiyear learning progressions)
Funding allocation	• Assessment vendors and technical assistance partners who support test development, psychometrics, test maintenance, etc.	• Assessment vendors and technical assistance partners • Teachers, LEA, and SEA capacity • Substantial proportions of the resources are more directly supporting teaching and learning rather than only test-related activities
Scoring	• Prioritize machine-scored items • Teacher involvement is a nice-to-have, not a must-have. Teachers may be invited to participate in some content development and review activities, but teacher involvement is not central to the goals of the assessment.	• Prioritize teacher-scored tasks alongside machine-scored items • Teachers are an integral part of assessment development and scoring because participation in these activities builds educator understanding of standards and the kinds of teaching and learning that are needed to excel in the discipline (and on the assessment).
Student experience	• Students experience the test as completely separate from instruction; they see it as an external judgment of their abilities in the tested area.	• Students experience the assessment as coherent with their instructional experiences, and as a meaningful experience in its own right. The assessment is engaging, motivating, and connected to what students have been asked to engage with in the classroom.
Implications for test design	• Prioritize selected-response items that can be developed and scored quickly and at low cost • Prioritize coverage; focus on the easiest-to-assess content • Acceptable trade-off that teachers, students, families do not see the assessment items and only see score reports	• Include tasks that model instructional shifts and sophisticated performance expectations • Prioritize surfacing information that signals and provides information about the most important instructional shifts, including disciplinary practices, disciplinary sense-making, and application of concepts and practices to relevant situations • Teachers must have access to at least a subset of the assessment tasks to build practice and capacity

Source: Learning Policy Institute. (2024).

Performance Assessments as a Path to Center Assessment as Learning in Large-Scale Systems

Performance assessment is an approach to assessment that measures students' knowledge, skills, and abilities by asking students to demonstrate their learning within real-world contexts that require it. Performance assessments might look like short, on-demand tasks, such as responses to an open-ended prompt; experiences that take up to a few hours, such as laboratory investigations, writing a paper, or presenting ideas; or extended, in-depth projects that happen over days to months and involve iteration, feedback, and revision, such as conducting and presenting original research or engaging in multidisciplinary projects.

What makes performance assessment a particularly compelling form of assessment is the authenticity of the tasks that students are asked to complete.[4] Performance assessments measure what students know and can do by asking them to apply their learning under the actual conditions under which that learning has relevance and value—closely approximating real-world applications and performances.[5] As a result, performance assessments accomplish the following three objectives:

1. **They emphasize depth of learning outlined in state standards by surfacing evidence of deep conceptual understanding, disciplinary practices, and higher-order thinking and problem-solving skills.** Current state standards in math, English language arts (ELA) and literacy, and science intentionally emphasize the kinds of transferrable disciplinary understanding students need to be successful in college and careers. This includes engagement in disciplinary practices and the application of content and practice in service of authentic problem-solving and critical thinking. While multiple choice and other selected-response item types can provide an efficient way to sample the breadth of standards and discrete pieces of knowledge, performance assessments are better suited to measure how well students are meeting the depth that state standards expect, along with elements like disciplinary practices that support learning transfer and are often considered most relevant to students' postsecondary success. Performance assessments can strategically sample multiple, specific standards in ways that attend to their disciplinary concepts and modes of inquiry through authentic application. In the absence of performance assessments, these aspects of standards are often under-assessed or ignored altogether, limiting the ways many large-scale assessment scores can (or should) be interpreted and used.[6]

2. **They directly support teaching and learning while yielding information about student performance in aggregate.** Performance assessments influence teaching and learning in the following important ways:

    - **Signaling.** Rich performance tasks provide concrete examples—and direct experiences—of the kinds of learning and performance opportunities students should have, signaling to educators and leaders the kinds of instruction needed to prepare all students to be able to successfully complete performance tasks.[7]

    - **Curriculum Equity.** Systems that integrate high-quality, curriculum-embedded performance assessments build curriculum equity by increasing access to sophisticated learning experiences for all impacted students.[8] They do so by offering all students who engage with the assessment a meaningful, engaging, and relevant experience through the tasks themselves; providing educative

experiences that build teachers' capacity for high-quality teaching and learning as they implement the tasks, evaluate student work, and determine how their curriculum may shift in the future; and giving leaders examples of what high-quality learning and performance should look like, supporting more effective decisions about curriculum materials, professional learning, and other systemic supports for teaching and learning.

- **More Useful Data and Scores.** Performance assessments are also incredibly information-rich: They simultaneously surface information about *what* students know and information about *how* they are thinking. When teachers look at student responses, they can see what students understand and can do and how they approach complex tasks requiring higher-order thinking skills. This kind of precise information is invaluable to teachers, students, and families who are interested in how to use student assessment data to improve instruction and outcomes for learners.[9] When designed appropriately, these same tasks can yield scores that can be aggregated and compared across student groups to inform decisions at the school, district, and state levels, scored in ways that are similar to constructed and extended response items and essays that are common on large-scale, high-stakes assessments.[10]

3. **They reflect parent, family, and community priorities.** Students and their families routinely criticize assessments for being irrelevant—noting that they do not engage students in addressing authentic situations and problems that matter and that they do not measure student learning relative to goals that they care about, like problem-solving, critical thinking, and the knowledge and skills needed to be successful in their next steps.[11] Rich performance tasks can position large-scale assessments to both engage students and measure learning meaningfully by asking students to use the knowledge and skills they have developed to make sense of an authentic problem, text, or real-world observations.

Defining Performance Assessment

"Performance assessment" refers to the spectrum of opportunities for students to demonstrate and receive feedback on what they know and are able to do through demonstrations of their knowledge, skills, and abilities in meaningful contexts that, together, more closely approximate how student learning is actually used in real-world settings.

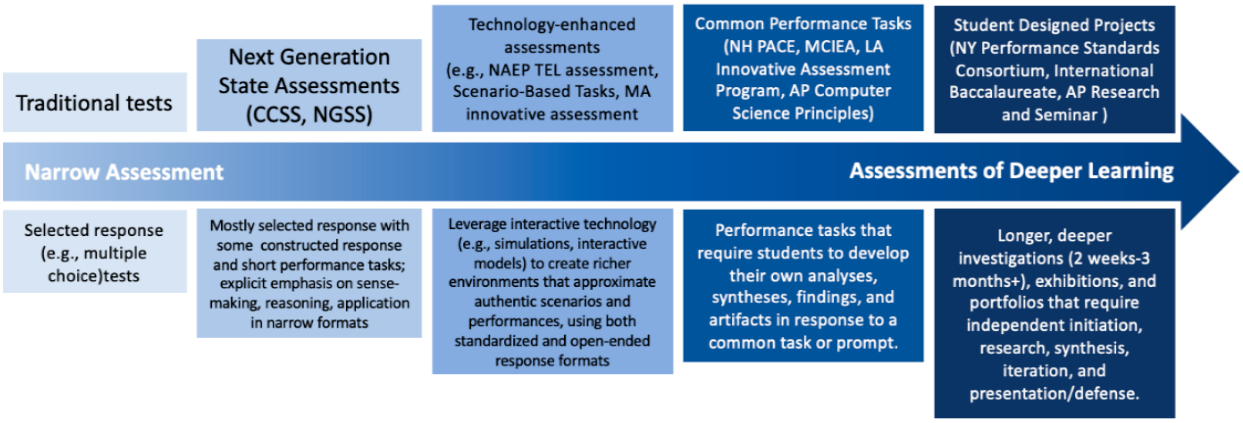
How are performance assessments different from traditional tests? They differ in what is evaluated: student artifacts rather than selected responses. Performance assessments provide an opportunity to measure and evaluate student understanding based on student responses to open-ended tasks that require original student artifacts that make their thinking visible. This might include constructing an argument, designing solutions, performing an activity, etc. This contrasts with assessment approaches that rely heavily on selected responses (e.g., multiple choice), which tend to reduce and decontextualize student performance, leading to scores and grades that are less indicative of actual student understanding and abilities.

How are performance assessments different from learning experiences or activities? Performance assessments are distinguished from instruction because they are designed to support evaluation, measurement, and feedback. They include clear evaluation criteria relative to specific targets

and inferences and are designed to surface empirical data on what students know and can do that can be used for a variety of purposes related to monitoring and supporting student progress. Importantly, performance assessments *can* be learning opportunities unto themselves and provide ways for students to develop their knowledge and skills further while they are demonstrating their current abilities. In deep learning systems where student inquiry drives instructional approaches, performance assessments often integrate seamlessly into instruction from both student and teacher lenses. (See Figure 1.)

Source: Darling-Hammond, L., & Adamson, F. (2014). *Beyond the bubble test: How performance assessments support 21st century learning.*

Figure 1. Assessment Continuum



Source: Adapted from Darling-Hammond, L. (2017). *Developing and measuring higher order skills: Models for state performance assessment systems*. Council of Chief State School Officers.

While performance assessments can evoke ideas about extremely individualized projects and performances that are too personalized to provide generalizable and comparable information about student performance, some performance assessments are designed for comparability and used as highly trustworthy measures of student performance in local, state, and national systems. When performance assessments are used to generate scores for local, state, or national purposes, they often involve common tasks or sets of calibrated tasks designed for common rubrics that can be reliably used and scored across classrooms and contexts, allowing assessments to generate comparable evidence of student performance and enabling reliable scoring. Evidence from states and countries that employ performance assessments suggests that the systematic and well-supported use[12] of performance tasks can serve to do the following:

- improve both teaching and learning,
- create curriculum equity by giving all students taking the assessment access to sophisticated and motivating experiences, and
- provide important data about student progress (individually, as well as at the school and district level) toward the depth and breadth of state standards to inform resource allocation and growth strategies.[13]

Designing High-Quality Performance Assessments for Systemic Use

Many large-scale systems—such as in states or districts, or national and international curricular and assessment programs like Advanced Placement (AP) and International Baccalaureate (IB)—use performance tasks as part of their system[14] because they elicit evidence of student thinking that is not readily surfaced through selected-response items and serve to influence instruction in positive ways. When designing performance tasks for use in larger-scale systems, developers and leaders often emphasize certain features of assessment design, implementation, and scoring of performance assessments. These features include common tasks and rubrics, calibrated scoring, and rigorous assessment development processes to produce tasks that contribute to trustworthy student scores that can be aggregated and compared as needed. These features distinguish performance tasks that can be used to generate trustworthy and comparable scores from those that are often developed locally as part of meaningful instruction but should not be used within large-scale systems (e.g., a project or task developed by an individual teacher to be used as part of coursework).

Common Tasks and Rubrics

Nearly all formal assessment systems that employ performance assessments ask students to complete *common* tasks. This approach could include:

- asking all students to engage in the same task;
- asking students to engage in one of a set of tasks intentionally designed to measure the same knowledge and skills, with the specific task chosen by educators, by students, or randomly assigned;
- allowing educators to leverage tasks that are already embedded in their curriculum and that are aligned to content standards and designed for criteria for assessment design; or
- providing educators and students with options for equated high-quality performance assessments, intentionally designed to connect with content and questions in the statewide summative assessment, to be administered as part of instruction.

In some assessment systems, tasks are designed such that local educators can choose when to administer these tasks in their classrooms, based on their particular curriculum choices. In other systems, all students engage in common performance tasks during the same testing window, similar to how other standardized assessments (e.g., state summative assessments) are generally administered. In many instances, common curriculum-embedded performance tasks are used in conjunction with a more traditional on-demand assessment to balance depth, breadth, and time spent on the assessment. In these cases, the performance tasks and their administration are standardized enough that they can be used reliably to generate comparable student scores. In many ways, formal assessment systems that use curriculum-embedded performance tasks treat these tasks as simply another item type—one designed to surface better information about student thinking and performance while being held to similarly high standards for technical quality. See Appendix A for examples of these kinds of common assessment tasks employed in large-scale systems.

Rubrics and Calibrated Scoring

Authentic performance tasks ask students to individually (and sometimes collaboratively) generate work in the form of essays, lab reports, math solutions, and the like. Consistent and reliable scoring processes depend on two equally important elements: clear, evidence-based rubrics and trustworthy, calibrated ratings among scorers.

- **Evidence-Based, Expert-Designed Scoring Rubrics.** Large-scale programs like AP, IB, the National Assessment of Educational Progress (NAEP), and state assessments develop rubrics for performance tasks based on content-specific research on how students learn; expert input, including expert responses to performance task prompts; task-based student work and cognitive labs; and the score interpretations a performance assessment is intended to support. These rubrics are carefully constructed, field-tested, and validated based on a wide range of student work and, in some cases, the statistical distribution of performance. The net result of these efforts is rubrics that are clear, discerning, and used to provide valid scores for student performance on open-ended tasks. Many programs use rubrics that are developed for competencies or learning goals (e.g., a rubric for investigation design in the life sciences, rather than a rubric for conducting an investigation of tomato plant growth) so that the same rubric may be used across many tasks, affording the possibility of a higher degree of comparability across different tasks.[15] In some programs, these more general rubrics are adapted to the specifics of a given task to aid reviewers in their ratings.

- **Scoring Approaches.** Reliable scoring can be accomplished in a number of ways, all of which depend on effective rating of student responses against well-designed rubrics. States, districts, and programs that use performance tasks invest heavily in training raters to ensure that they develop a shared conceptual understanding of the assessment target, a common interpretation of rubric-specified criteria for student performance, unbiased interpretation of evidence of student performance in student work, and consistency across raters in terms of how they apply each rubric to make judgments about student work. Some programs rely on raters expert in content and pedagogy to build the training upon a strong conceptual foundation. In other circumstances, programs focus on extensive calibration across panels of trained educators such that rating student responses can be distributed across a wider range of raters and can contribute to meaningful professional learning for teachers. In cases where centrally trained experts and educators are scoring student work, multiple raters will often score each piece of work, either developing a consensus score, calculating interrater reliability, or both. Some school systems allow trained local teachers to score student work and employ a central audit of a certain representative sample of scored work to ensure consistency and interrater reliability. In nearly all cases, these processes make use of expert-vetted student work benchmarks to anchor scorer training and calibration as well as validation of scoring itself.

Rigorous Assessment Development Processes

Developing high-quality performance assessments is an intensive process.[16] Before tasks are even outlined, the development process involves unpacking content and performance standards (e.g., state standards, 21st-century skills, and competencies) to understand what knowledge, abilities, or performances are to be measured by assessments and what kind of evidence of student performance tasks must elicit to determine the degree to which students have progressed toward those targets. Once developers have a clear understanding of student learning expectations and the specific targets for assessment tasks, they develop criteria and features of tasks designed for those targets, specifying the kinds of stimuli and prompts that need to be given, how to surface thinking at appropriate levels of sophistication, and what knowledge and skills are within and out of bounds for the assessment. When assessment tasks are drafted to meet these criteria, they are reviewed by educators and field-tested with educators and a diverse set of students to ensure that the tasks are functioning the way they were designed. Final decisions about including performance tasks within assessment systems involve considering alignment to standards and item or task performance statistics, as well as evaluation of student work, data from cognitive labs, and observations of students and teachers.

The development process involves unpacking content and performance standards to understand what knowledge, abilities, or performances are to be measured.

Tuning Performance Assessments Based on Purpose

Performance assessments can be powerful tools for transforming teaching and learning in K–12 systems. One of the strengths of taking a performance-based approach to assessment is that performance assessments can be designed to surface evidence of a wide range of student knowledge, skills, and abilities for a number of purposes. For example, a student portfolio and defense might be appropriate as part of the case for student readiness for graduation or advancement, but it might not be as useful as a stand-alone curriculum-embedded task for assessing progress at a particular point in time. Given the range of purposes and forms that performance assessments can take, it can be helpful to think about performance assessments as being governed by a set of *constants* (see Figure 2) that distinguish high-quality performance tasks from other kinds of assessment and a series of *knobs* (see Figure 3) that can be turned depending on the purposes and intended uses of the assessment. Constants and knobs can be understood as follows:

- **Constants:** common features that are true of all high-quality performance assessments. Effectively, these are what characterize, or set apart, performance assessments. Constants represent the "non-negotiables" of high-quality performance assessments. These features ensure that performance assessments are trustworthy measures of student learning.
- **Knobs:** variable features to which performance assessments can—but do not always—attend. A given high-quality performance assessment may foreground some knobs and background others, and no single assessment task will equally attend to all knobs. Over time, performance assessments can be designed such that a complete picture of student learning is surfaced across the system.

These features help to make performance assessments valid and useful measures of student learning. Together, these common and variable features provide a framework for understanding how to design, use, and interpret performance assessments within K–12 systems.

Figure 2. Constants or "Non-Negotiables" for Performance Assessments

Source: Learning Policy Institute. (2024).

Figure 3. Knobs to Adjust for Performance Assessments

Source: Learning Policy Institute. (2024).

Constants and Knobs Used to Tune Performance Assessments Based on Purpose

Constants

Across different assessment purposes, all high-quality performance assessments include common features such as the following:

- **Designed for Appropriate, Targeted, and High Expectations.** Performance assessments are designed for and aligned to the standards, expectations, targets, and competencies they are designed to assess, in accordance with best practices of the discipline(s). Importantly, high-quality performance assessments are designed to assess student performance relative to high expectations. They provide an opportunity to lift the ceiling on student performance, rather than lower the floor.
- **Are Open-Ended and Generative.** Performance assessments ask students to create artifacts that make facets of student thinking (including targeted knowledge, skills, and abilities [KSAs]) visible at a range of levels; are evaluated relative to the targeted expectations, in order to successfully respond to the assessment task; and focus on activities and behaviors that center sense-making and meaning-making.
- **Are Equitable, Fair, and Accessible to All Students.** All high-quality performance assessments meet common, baseline requirements for equitable and fair assessments: They are accessible, free from bias, and include supports that enable diverse learners to demonstrate what they *do know* and what they *can do*. Moreover, many high-quality performance assessments also intentionally provide avenues for student agency and identities to be lifted up as an important and valued component of engaging in the assessments in ways that are appropriate to the format, purpose, and scope of the assessment task.
- **Require Authentic Student Sense-Making.** Performance assessments ask students to engage in activities that are authentic (to the world and to the discipline[s] being assessed), rigorous, and require students to make their KSAs visible through the process of making sense of phenomena and problems. Well-designed performance tasks ensure that students' diverse experiences are valued components of the sense-making shown.
- **Include Clear Evaluation Criteria.** These clear criteria (e.g., rubrics, scoring guidance) reflect the expectations assessed and specific inferences to be made. Performance assessment tasks are designed to help appropriate interest holders (including students and other users of the information) make specific claims about student performance that are aligned to the targeted expectations. Evaluation processes, criteria, and reports (e.g., rubrics, scores, achievement or performance level descriptors) specifically connect to an appropriate range of student work to inform those claims.

Knobs

In addition to these common features, performance assessments can be designed to highlight, support, and measure particular additional competencies that are essential for college, career, and civic readiness. While no single performance task will measure all of these features equally, a series of performance tasks can together, over time, reveal a complete picture of student learning. These variable features include the following:

- **Student Choice/Agency.** The degree to which students have ownership over the why, what, and how of the assessment task.
- **Connection to Student Identities.** The degree to which students' cultural and intellectual conceptualizations of themselves are engaged as an important part of the task.
- **Disciplinary Sophistication of the Performance.** The degree to which the task engages students in complexity, nuance, and uncertainty authentic to the discipline(s) the task is situated within.
- **Authenticity to the Real World.** How real the problem, phenomenon, and task is, including whether an authentic audience is included during students' development of evidence that meets the asks of the task.
- **Transfer.** The degree to which tasks require the application of KSAs in contexts that are different than the learning context(s).
- **Creativity.** The level of non-routine sense-making that tasks allow or require, building from a foundational understanding of content knowledge.
- **Collaboration.** The degree to which artifacts produced are the result of and indicate effective collaboration.
- **Integration Across Domains.** The degree to which multiple content areas (within or across disciplines) are integrated in the student work to be generated.
- **Student Reflection and Metacognition.** The degree to which student reflection and metacognition on their own progress, growth, and performance is an explicitly important part of the task (e.g., opportunities are offered for self-assessment, revision, subsequent goal-setting).
- **Range of Evidence Generated and Used.** The quantity and types of artifacts and the time period over which they are generated (e.g., portfolios or individual tasks, iterative artifacts developed in response to feedback).
- **Coverage.** The depth and breadth of targeted expectations included in an assessment, relative to the full range of student expectations (e.g., given the full set of KSAs and dispositions targeted for a unit or course, what proportion of KSAs are students asked to demonstrate on a given assessment task/assessment instrument, such as a final project or exam?).

Source: Learning Policy Institute. (2024).

Performance Tasks in Action in Large-Scale Programs

Many large-scale programs already incorporate sophisticated performance tasks as part of their current assessment designs. By examining why and how these systems currently use performance assessments, we can better understand how performance assessments can be designed to contribute to higher-quality assessments that intentionally support meaningful teaching and learning. Some examples of large-scale programs using performance tasks are the following:

- **Smarter Balanced Assessment Consortium.** In Smarter Balanced assessments, students complete short performance tasks in English language arts (ELA) and math as a significant part of their summative test score. These tasks, defined by the fact that they do not include selected-response items, ask students to integrate their understanding of content and practice across multiple standards in service of authentic scenarios and problems. The ELA test includes a 2-period research and writing task in which students have to bring evidence to bear in making a case about a particular problem. The math test includes a 1-period task in which students need to solve a problem, such as determining where to go on a field trip given the costs of different options, or how to design a garden given the size of the space and the needs of different kinds of plants. These tasks are completed independently by students, and student work is human-scored against validated rubrics. These scores then contribute score points to the overall raw and scaled scores for student performance.

- **Massachusetts Innovative Science Assessment.** Under the U.S. Department of Education's Innovative Assessment Demonstration Authority (IADA), Massachusetts is developing an innovative summative assessment in science that relies on simulation-based on-demand performance tasks. These individual performance tasks, in conjunction with a shortened version of the state's current science assessment (Massachusetts Comprehensive Assessment System—MCAS Science), will be used to generate individual student scores that can be used to meet current federal requirements for assessments. The state is currently exploring complementing its innovative summative assessment with curriculum-embedded performance tasks that include collaborative elements as well as individual student work and allow concrete feedback for high-quality teaching and learning. Both the on-demand simulation-based performance tasks, as well as the extended curriculum-embedded performance tasks, leverage carefully developed common assessment tasks to better align to the state's science standards and provide more valid and useful data than the current summative assessment alone.

- **Performance Assessments in Advanced Placement (AP) Assessments.** Several AP course assessments—used across a wide range of curricular, demographic, and geographic contexts—use performance assessments as a significant component of students' final scores. For example, AP Computer Science Principles (CSP) requires that students create a computer program during class that solves a problem, supports an innovation, explores a personal interest, or expresses creativity. This requirement includes both individual and collaborative components: Students are allowed to collaborate with peers on the development of the program and must submit an individual written description and video. These performance tasks are centrally scored by experienced, trained, and calibrated educator reviewers—just like other open-ended AP exam items—and contribute to students' overall course scores, which can then be used by colleges and universities for course

credit. AP CSP leverages the authenticity of the task and clear guidelines for the computer program to ensure that the task can be completed in different ways by students and still generate comparable student scores.

Similarly, the AP Research and AP Seminar courses require students to submit a digital portfolio of multiple curriculum-embedded performance tasks (e.g., individual research report, individual and team presentations, oral defense, evidence-based arguments). These tasks are scored by a combination of students' own teachers and centralized expert educators, all of whom have been trained and certified to implement and evaluate the performance tasks. Like AP CSP, AP Research and AP Seminar rely on authentic tasks with clear evaluation criteria, allowing for flexible but comparable teaching, learning, and student performance. The College Board is currently exploring including performance assessments on a much wider range of AP courses, using similar design, implementation, and scoring approaches.

- **Performance Assessments in International Baccalaureate (IB) Courses.** Students complete a wide range of assessment tasks that are performance based, including open-ended essay examinations, portfolios of work, projects, and research investigations. These are scored by trained teachers using rubrics or other standardized tools for evaluation in a process that is moderated to ensure that judgments are reliable. A hallmark of IB assessment is that it strives always to be authentic, placing a high value on testing what is important in a way that reflects the real world.

- **New York Performance Standards Consortium.** In the New York Performance Standards Consortium, graduating students complete Performance-Based Assessment Tasks (PBATs) including an analytic essay on literature, an original scientific investigation, problem-solving in higher-level math, and an original social studies research paper. These tasks are developed as part of students' ongoing curriculum-embedded experiences, and they are evaluated based on rigorously developed common disciplinary rubrics used for both written presentations and oral defenses of their work. The development and calibration processes for the rubrics ensure scoring reliability across learning contexts and projects. Research suggests that because of the Consortium's sophisticated and comprehensive support for performance assessment throughout students' high school experiences, student performance on graduation PBATs can be effectively used to support college admissions and success, serving as an indicator of their validity.[17]

- **New York State Regents Exams.** New York's Regents Exams include common, on-demand performance task components for a range of subject areas that are used as part of students' summative scores. For example, students develop written text-based arguments on ELA assessments, respond to document-based questions on history assessments, and develop explanations for phenomena and problems on science assessments. These open-ended tasks are collaboratively developed by educators and assessment development staff and are scored by teachers who receive training and use detailed scoring rubrics. Scores are audited and validated by the state.

Curriculum-Anchored Assessments Across the State Performance Assessment Learning Community

A growing number of states are seeking to establish performance assessment systems that are anchored in the learning progressions and activities embedded within high-quality curriculum. With the rise of states establishing policies and investments that promote high-quality instructional materials and professional learning, leaders have a unique opportunity to build assessment systems that are coherent with teaching and learning policies that can, together, advance deeper learning more effectively than through assessment and accountability approaches alone.

While the specific assessment strategies that states are using vary by context, nearly all states pursuing this model are making available high-quality and common curriculum-embedded performance tasks and rubrics that are intentionally designed to be coherent with a range of high-quality and vetted curriculum materials. In some cases, state strategies focus on supporting the implementation and use of these curriculum-embedded tasks through:

- local performance assessment requirements or state-provided interim assessment systems;
- statewide scoring and interpretation support;
- guidance for principled adaptation of a range of high-quality curricula to promote effective instructional shifts, based on student work analysis; and
- development of a vetted library of curriculum-anchored assessment tasks that are tagged to different curricula, with instructional materials–specific guidance for timing, use, and interpretation.

In other cases, states are seeking to directly incorporate student performance on these tasks into summative assessment scores. In addition to considering how to incorporate scores on the performance task directly into student summative scores, many states are considering a direct link between curriculum-based activities and the on-demand summative assessment as a way for curriculum-embedded performance tasks to have consequential impact on student scores. This might look like: (a) including summative assessment items that ask students to reflect on data/outcomes from a curriculum-embedded performance task, (b) including phenomena or problems students explored in curriculum-embedded settings, or (c) leveraging the kinds of simulations and technology students experience on a curriculum-embedded performance task within summative assessments. This approach allows the tasks used to be relatively flexible and transparent—teachers can decide when and how to administer the assessments and can immediately see how students performed on the task to inform next steps in instruction—while still counting student engagement in sophisticated performances as part of their overall summative scores.

Source: Learning Policy Institute. (2024).

Another approach to performance assessments integrates them as a central element of instructional materials and curricular experiences, graduation pathways, and resources made available systemically to teachers. These include:

- **Inquiry Units and High-Quality Instructional Materials.** One primary purpose of performance assessment is to drive high-quality, inquiry-oriented teaching and learning. Instructional units that center student-driven approaches can include embedded performance assessments, such that assessment is completely coherent with students' learning experiences. In inquiry units that leverage embedded performance assessments, the assessments themselves advance progress along learning progressions by simultaneously providing learning opportunities and feedback that is connected to concrete and actionable next steps for teachers and students. For example, OpenSciEd instructional materials are designed to engage students in making sense of science phenomena as the primary mechanism to develop their understanding of science concepts and practices. Embedded within each inquiry unit are robust formative and summative tasks that ask students to plan and carry out investigations, develop and revise models, and engage in argumentation to hone explanations for observations that they make. From the student perspective, these assessment opportunities are seamless with learning; they are the part of the process of figuring out why something happens. From the teacher perspective, these embedded assessments include clear guidance for how to interpret ranges of student responses, what feedback to share with students, and what next steps should be taken as part of daily instruction.

  Similarly, the Multiple Literacies in Project-Based Learning (ML-PBL) approach to teaching science in elementary school engages students in inquiry units and embedded performance assessments that integrate science, math, ELA, and literacy and emphasize social-emotional learning. Students progress toward science, math, and ELA standards through collaborative investigations of questions like "Why do I see so many squirrels but can't find any stegosauruses?" An assessment in ML-PBL might look like creating a short story to share a model of environmental change based on fossil evidence and students' understanding of environmental impact on living things. As with OpenSciEd, ML-PBL assessments offer students an engaging and meaningful learning experience as well as a way to demonstrate what they have mastered to date. For teachers, they open a window into cross-disciplinary competencies as well as into student ability to work with others, iterate and respond to new information, and creatively problem-solve. The materials are designed to provide just-in-time information linked to next steps within the learning sequence. Research on curriculum efficacy showed that students engaged with ML-PBL outperformed their peers on statewide science assessment measures, and students more frequently noted the value of collaboration and reflection on their learning.[18]

- **Common Performance Tasks.** Another effective way to build performance assessments into K–12 systems is through the use of common performance tasks that can be used across curricula, classrooms, and instructional contexts. For example, the Balanced Assessment in Mathematics includes a series of carefully designed, curriculum-agnostic performance tasks designed for math learning progressions, along with instructional supports for interpretation and next steps that teachers can use as appropriate in their classrooms.

Common performance tasks can also be useful components of statewide assessment systems, used either in tandem with selected-response items, such as the essay questions or constructed-/extended-response items included on many large-scale assessments, or as a replacement for some end-of-instruction tests. For example, in New Hampshire, the Performance Assessment of Competency Education innovative assessment and accountability pilot uses local performance tasks and common performance tasks across districts in lieu of a portion of math, ELA, and science end-of-year assessments. These common tasks are designed to be embedded directly as part of classroom activities, and are validated with limited, periodic standardized assessments. The use of common assessment tasks allows for comparisons to be made across classrooms and schools, while supporting more meaningful teaching and learning.

- **Capstone Projects and Tasks.** Inquiry units and common performance tasks lean into the value of common performance-based approaches to teaching and learning for all students engaged with those activities. Capstone projects offer students the opportunity to engage in rigorous inquiries (e.g., research projects, extended papers, community-based projects) that are individualized to students, allowing them to pursue a topic or question of personal relevance. Capstone projects are often evaluated using common rubrics that include targets for disciplinary learning as well as demonstration of 21st-century skills, allowing these experiences to contribute to students' demonstration of mastery and readiness for next steps. For example, in Oakland Unified School District, students complete an original research-based capstone project as part of their graduation requirements, intended to demonstrate student competencies aligned to the district's profile of a graduate. Students pursue projects that have meaning to them, and these projects are often community based and civically minded. To support evaluation of these individualized efforts, the district uses common rubrics for writing, oral presentation, and field research. The emphasis on the common higher-order skills that are needed across inter- and cross-disciplinary projects allows students to pursue different topics with a common progress- and mastery-monitoring system in place.

- **Student Portfolios and Defense.** Portfolios of learning allow students to curate a set of evidence over time and across performance assessment opportunities to demonstrate competencies/standards. Portfolios leverage tenets of performance assessment in two ways. First, curriculum-embedded performance assessments, such as papers and project artifacts, are often the basis for the evidence students collect as demonstration of the targeted competencies. Second, the act of curating and defending the portfolio—often requiring students to demonstrate features like growth and problem-solving—is a performance assessment that asks students to self-reflect and engage deeply with their learning journey and their understanding of the competencies they are seeking to show. For example, in Los Angeles Unified School District, students collect artifacts and evidence of their content knowledge and growth while in high school, reflect on those artifacts, and present this evidence along with a reflection to a panel of peers, educators, and community members.[19] This panel determines each student's readiness to graduate from high school using a shared rubric to assess the quality of the student's artifacts and presentation. Artifacts may be drawn from coursework within a class or across classes and years, internships associated with their graduation pathway, and other activities outside of formal schoolwork that can be used as evidence for mastery of learning targets (e.g., set design for a school play).

Recommendations for System Leaders

Performance assessments can transform assessment systems into forces for improved teaching and learning. Doing so requires that system leaders position performance assessments—and the supports needed for their design and use—as a valued element of both instruction and student performance. As leaders consider how to reorient their assessment systems toward instructional relevance, they can keep in mind the following recommendations:

Performance assessments can transform assessment systems into forces for improved teaching and learning.

- **Demand assessments that measure what matters.** Ensure that assessments actually measure the higher-order thinking and problem-solving, disciplinary practices, and other deeper learning competencies that are needed for students to be ready for college, careers, and citizenship. Too often, large-scale assessment systems are designed with easy scoring in mind, leading to multiple choice assessments that are readily scored with no room for interpretation. However, prioritizing the ease of scoring over scores that actually reflect important aspects of students' knowledge, skills, and abilities (KSAs) leads to assessments that teachers, students, and families cannot use to meaningfully guide student learning. At best, this renders assessment systems irrelevant to stakeholders they are meant to support; at worst, this leads to assessment systems that drive educators and leaders to pursue ineffective strategies in an effort to chase an inaccurate representation of student success.

- **Recognize the transformative potential of signaling and student experience.** High-quality performance assessments will almost certainly ensure that student scores are more reflective of what students know and can do in the tested discipline. Important as this is, it may not be the biggest value-add of performance assessments within formal assessment systems. Large-scale assessment systems, which are often complementary to more instructionally sensitive measures that happen on an ongoing basis in the classroom, frequently make their biggest mark on instruction through their signaling function, influencing decisions about what gets taught, how students experience learning, and what success should look like. Including performance assessments in assessment systems can be transformative because doing so encourages instructional shifts toward deeper learning. Traditional large-scale assessments that rely heavily on selected-response items can limit instruction by narrowing curricula and incentivizing school systems to prioritize short-term regurgitation of facts and procedures rather than the deep engagement that leads to sustained learning and retention.

- **Leverage performance assessments strategically.** Many large-scale systems that leverage performance assessments do so in conjunction with other assessment instruments, such as on-demand selected-response items. The combination of both approaches to assessment allows assessment designs that can sample wider coverage while still providing substantial and sufficient evidence of students' ability to reason in sophisticated ways within and (if appropriate) across disciplines. The key is striking a strategic and effective balance by ensuring performance assessments are communicated and counted for enough of students' final scores that stakeholders

pay attention to the KSAs needed to complete performance tasks on a given assessment, while providing space for other assessment items to measure sufficient breadth. Consider emphasizing the role of performance assessments by designing assessment blueprints that allocate sufficient time and score points to meaningful performance tasks, releasing high-quality performance tasks and anonymous student responses reflecting various levels of performance as part of released and annotated items, and prioritizing the importance and role of performance tasks in system-led communication and professional learning opportunities.

- **Make the assessment worth the investment.** A major element of the value proposition of performance assessments is their authenticity and educative nature. Leaders should consider leaning into the assets of performance-based assessments by prioritizing the development and use of authentic, relevant, and sophisticated tasks that motivate students and provide a beacon of what their routine experiences in the classroom should look like. Leaders also should engage all classroom teachers in the development and interpretation of these tasks so that teachers can have access to the rich information about student thinking that such tasks produce and to build support for the pedagogy that enables deep learning.

  When performance assessment systems have failed to drive lasting change, they have done so not because they were not feasible to design and implement, but because they were implemented in ways that did not maximize what these kinds of assessments offer. Many large-scale systems sacrifice the assets of performance assessments—measuring deeper learning, insight into student thinking, relevance and authenticity—in the name of test security and easy scoring. When they do so, system managers eventually have to ask educators, leaders, and even administrative staff to commit more time and resources to implementing the assessments while producing very similar information to traditional assessments dominated by selected-response items—often in the form of decontextualized scores. When performance assessments are not positioned to provide further insight about student thinking, it can be difficult for educators and leaders to see why performance assessments are worth the investment.

- **Consider creative resource allocation.** Many systems that center performance assessments emphasize the impact on teaching and learning for students, and also recognize the potential for meaningful and sustained professional learning, when making resource allocation and budgeting decisions. While performance assessments cost more to design and score than multiple choice questions, resources can be reallocated from test preparation and interim assessments that are essentially practice tests. Professional development time can also be allocated for design and scoring, as teachers consistently note the benefits of designing and reviewing performance assessments for their own learning and planning.[20] Resources spent on meaningful performance task development and scoring can contribute to better teacher professional development *and* student learning experiences—which is likely to lead to better student outcomes than simply practicing the questions on a superficial final exam.

Conclusion

As a growing number of states and systems consider how to break the habit of assessments being used to limit learning opportunities for students, leaders are considering a different purpose—and a different set of trade-offs—for their assessment system designs. Many leaders are drawing a line in the sand, saying that teaching and learning are paramount, and any assessment system that does not have a positive impact on teaching and learning cannot be acceptable. When leaders make positive instructional impact a necessary condition of high-quality assessment systems, performance assessments routinely emerge as an important element of system designs. Done well, and in conjunction with other measures, performance assessments can provide better evidence of what students know and can do while helping students and teachers alike better understand how meaningful instruction should look and feel.

As with any system, the devil is in the details. Designing instructionally relevant performance assessments that can be meaningfully used at scale requires keeping calibration, consistency, and quality in mind as core design features. Doing so—by focusing on rigorous development processes as well as rubric development and scoring calibration—can help transform assessment systems from a "necessary evil" or an ineffective postmortem analysis of instruction to a true partner in learning.

Appendix A: Performance Assessment System Details and Sample Tasks

This appendix includes six examples of assessment systems that leverage performance tasks as a substantial component of the system. These examples are:

1. AP Computer Science Principles (see Table A1)
2. Massachusetts Innovative Science Assessment (see Table A2)
3. New York Performance Standards Consortium (see Table A3)
4. New York State Regents Exam (see Table A4)
5. Smarter Balanced Assessment Consortium Performance Tasks (see Table A5)
6. Washington Civics Assessment (see Table A6)

Table A1. AP Computer Science Principles

Advanced Placement Computer Science Principles	
What does the performance assessment look like?	Students develop an app/computer program to address a student-selected purpose based on an authentic problem or their individual and group interests. Students independently or collaboratively develop the program code, acknowledging collaborators. When students submit their app, their code is accompanied by an independently developed video submission that illustrates the program running, including inputs, functionality, and outputs. They also independently submit four written responses describing: (1) the purpose, functionality, inputs, and outputs of the program; and (2) analysis of elements of the program code (lists, algorithm, testing) that highlight specific knowledge skills related to computer science.
How is the performance assessment used?	This performance task is administered during the school year as part of classroom activities. It is uploaded to a digital portfolio and scored by expert educator raters. Student performance on the task is included as a significant portion (30%) of the calculation of the student's overall score (scale of 1–5), in conjunction with an on-demand assessment that comprises largely selected-response items.
How is the performance assessment developed?	The performance task is developed through a rigorous principled design approach and is designed around a set of criteria for high-quality performance, drawn from the standards for the course. By focusing on alignment of the assessment to these criteria and standards, the task is able to be flexible to attend to student interest and local curricula while generating comparable student data that can be used to contribute to the high-stakes scores.

Advanced Placement Computer Science Principles	
How is the performance assessment scored?	Each task is scored against a detailed rubric that includes the learning objectives assessed, six scoring criteria, and decision rules for making a determination about score points awarded. Each task is scored by expert AP raters who have been extensively trained on the use of AP CSP rubrics for the performance task.
What knobs does this task prioritize?	Student interest, disciplinary authenticity, disciplinary sophistication, collaboration, creativity.

Source: College Board. (2020). *AP computer science principles: Course and exam description*.

Table A2. Massachusetts Innovative Science Assessment

Massachusetts Innovative Science Assessment	
What does the performance assessment look like?	The innovative science assessment uses scenario-based simulations to approximate real-world scenarios for students to figure out. Students engage with the scenarios and simulations, answering a coherent set of questions that leverage creative item types, technology enhanced items, and open-ended responses. These tasks may ask students to design an experiment, use a simulation to conduct an investigation or test a design solution, evaluate claims or observations, etc. Massachusetts is also piloting a curriculum-embedded component that links the summative assessment with features of high-quality instruction more directly, including simulations and collaborative sense-making opportunities in integrated tasks.
How is the performance assessment used?	The simulation-based performance tasks comprise the bulk of students' on-demand summative assessment in the IADA pilot. Together with a shortened version of the existing science Massachusetts Comprehensive Assessment System or "existing science state assessments (MCAS), these performance tasks contribute to individual student scores.
How is the performance assessment developed?	Each task is developed using an evidence-centered design approach by the state's assessment vendor. Tasks are designed and reviewed for alignment to standards, bias and sensitivity concerns, language accessibility, and developmental appropriateness.
How is the performance assessment scored?	Tasks are scored centrally by expert scorers. Each task is scored against an answer key and scoring guide that includes the learning objectives assessed, scoring criteria, and narrative to support decisions about awarding each score point. Scoring guidance also includes sample student work at each scoring level to support normed scoring practices.

Massachusetts Innovative Science Assessment	
What knobs does this assessment prioritize?	Disciplinary sophistication, transfer.

In this classroom performance task, students are asked to consider the increase in concussions young people are experiencing in youth and professional soccer. The task asks them to make a recommendation about whether headers (when heads collide with the soccer ball) should be banned in the sport.

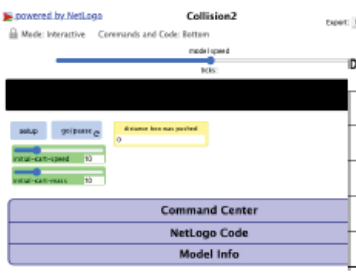
After reading short introductory material about the phenomenon and watching a short video, students are asked to complete three sets of activities:

Part 1: Individual Work

- Record initial ideas about how forces from a lightweight soccer ball could cause injury when it comes in contact with someone's head
- Develop an initial model of forces at work when a head collides with a soccer ball.
- Consider three types of head collisions that can happen in soccer, and consider how objects exert forces on each other in each situation through selected response, words, and drawings.

Part 2: Group investigations

- With a group of peers, students brainstorm possible factors and investigations that would determine how different kinds of head collisions would impact the likelihood of a resulting concussion.
- Based on their initial brainstorm, groups of students use a simulation to understand the **impact of mass and speed on forces involved in head collisions**.



Data Table:

Trial Number	Cart Speed (m/s)	Cart Mass (kg)	Distance Box was Pushed (m)
Trial 1			
Trial 2			
Trial 3			
Trial 4			

- They then conduct an investigation using a hands on model to understand how hardness of an object impacts the force of a collision, using weights, foam padding, and crackers in the classroom.
- With their group, they make sense of the finding from both the simulation and hands-on investigations. They develop models for how speed, mass, and hardness impact force.

Part 3: Individually putting the process together

- Students individually respond to two questions to synthesize and apply what they made sense of in their group investigations, transferring their understanding back to address the soccer situation:
  - Which type of collision (X, Y, or Z) do you predict would be most likely to cause a concussion? Use evidence from your group investigations and ideas you have learned in class.

  - Construct an argument that addresses the question: **Will banning headers in youth soccer significantly reduce the amount of concussions?**

Source: Massachusetts Department of Elementary and Secondary Education. (2024). *Science and technology/engineering assessment resources*.

Table A3. New York Performance Standards Consortium—Performance-Based Assessment Tasks

New York Performance Standards Consortium—Performance-Based Assessment Tasks (PBATs)	
What does the performance assessment look like?	Performance assessments are administered in each core content area and include a written paper and oral defense of an extended disciplinary activity (e.g., mathematical problem-solving, analytical essay, original science investigation, research paper). The tasks are flexible and designed in collaboration between students and teachers.
How is the performance assessment used?	The graduation-level PBATs are used to provide a summative evaluation of student performance. These PBATs are part of a comprehensive system that includes instructionally embedded performance assessments and disciplinary inquiry throughout classroom instruction, building toward these culminating performance tasks.
How is the performance assessment developed?	PBATs are collaboratively developed among teachers and students, intended to be responsive to student interests while maintaining rigor. Performance tasks grow out of meaningful work being pursued in the classroom and are subject to extensive vetting through cross-site moderation studies during which Consortium teachers evaluate tasks, rubrics, and student work to norm on task quality. Tasks are also externally vetted by evaluators and superintendents from each participating school district.

New York Performance Standards Consortium—Performance-Based Assessment Tasks (PBATs)	
How is the performance assessment scored?	Performance tasks—both written and oral components—are scored by local teachers and external evaluators using common, validated rubrics for each task type.
What knobs does this assessment prioritize?	Student interest, disciplinary authenticity, disciplinary sophistication, audience authenticity, integration across domains, transfer.

Mathematics PBAT (sample materials)

Introduction: general overview of the course (*Length: approximately a half page*)

Below are some prompts to help you think and write about what you know and understand about "proofs." Don't simply answer each question as if it were part of a questionnaire—instead, incorporate your response within your introduction in narrative form.

Possible ideas:

- What are proofs? Do we need proofs? When and how do we use proofs?
- Are proofs important in mathematics? What types of proofs are there? Etc.

Introduction of your specific PBAT topic: What are you going to prove in your paper?
(*Length: it may take less than half a page. It depends on the topic.*)

Provide some brief background information about your topic. State your claim (*what will you prove?*)

Proof (the actual work)
(*Length: It depends on the topic; diagrams and graphs will add pages to the paper.*)

Remember, "more" is not necessarily better; however, make sure to include all relevant information, visuals, tables, diagrams.
Examples:

- if you are doing a Number Theory proof, then write: "let's assume ***n*** is an even number, then we can express $n = 2k$ because …"
- if you are doing the Pythagorean theorem proof, then state what each variable stands for, etc.
- if you are doing a syllogism, indicate what each letter stands for. Don't assume anything, even if it appears obvious. State it!

Support each statement you make along the way with evidence.
The writing must be fluent, clear, and easy to read and understand. Refrain from repeating yourself, as it makes the paper confusing. The reader may lose interest in reading as well.

Conclusions

After summarizing your findings, state your final conclusion. Generally, you repeat the introduction of your PBA Task with the difference that now you are confirming it.

New York Performance Standards Consortium—Performance-Based Assessment Tasks (PBATs)

Performance Indicators	Outstanding	Good	Competent	Needs Revision
Problem Solving	Selects appropriate and efficient strategies to solve non-routine problems. Provides in-depth analysis of strategies Executes conceptually sound mathematical procedures accurately.	Selects appropriate and efficient strategies to solve non-routine problems. Provides some analysis of strategies Executes conceptually sound mathematical procedures with minor computational errors.	Selects appropriate, but inefficient, strategies to solve non-routine problems, and executes conceptually sound mathematical procedures with minor computational errors. or Selects appropriate and efficient strategies to solve non-routine problems but executes mathematical procedures with minor conceptual and computational errors.	Selects an inappropriate strategy or Makes major conceptual errors or procedural errors.
Reasoning & Proof	Makes valid conceptual/theoretical argument(s) and mathematically justifies it logically and thoroughly.	Makes valid conceptual/theoretical argument(s) and mathematically justifies it logically.	Makes argument(s) and justifies most mathematical statements accurately.	Makes arguments but does not justify mathematical statements accurately.
Communication	Always uses mathematical language and notations accurately. Always clearly explains mathematical thinking in an organized and detailed way.	Mostly uses mathematical language and notations accurately. Mostly clearly explains mathematical thinking in an organized and detailed way.	Sometimes uses mathematical language and notations accurately. Sometimes clearly explains mathematical thinking in an organized and detailed way.	Limited use of mathematical language and notation in an accurate manner. Rarely clearly explains mathematical thinking in an organized and detailed way.
Connections	Demonstrates an in-depth understanding of the relationships between mathematical concepts, procedures, and/or strategies.	Demonstrates an understanding of the relationships between mathematical concepts, procedures, and/or strategies.	Demonstrates a limited understanding of the relationships between mathematical concepts, procedures, and/or strategies.	Does not demonstrate understanding of the relationships between mathematical concepts, procedures, and/or strategies.
Representation	Creates an accurate and sophisticated mathematical representation(s), inherent to the task, to solve problems or portray solutions.	Creates an accurate mathematical representation(s), inherent to the task, to solve problems or portray solutions.	Creates an accurate mathematical representation(s), inherent to the task, to solve problems or portray solutions, but may be imprecise or contain minor errors.	Does not create an accurate mathematical representation, inherent to the task, to solve problems or portray solutions.

Source: Information about PBAT design and use, as well as libraries of rubrics and student work, can be found at New York Performance Standards Consortium (accessed 09/06/2024).

Table A4. New York State Regents Exams

New York State Regents Exams	
What does the performance assessment look like?	There are performance-based components of several Regents exams, including English, history, and science courses. In each case, the tasks are short, on-demand activities that ask students to engage with meaningful stimuli (texts, document-based sources, science problems and phenomena) and generate work to make sense of the stimulus. This may look like a written argument, scientific explanation, etc.
How is the performance assessment used?	The performance assessments are scored and used as part of students' summative assessment score.
How is the performance assessment developed?	Performance tasks are developed through a collaboration between the state education agency and educators. The tasks are developed using detailed assessment development routines that ensure tasks are designed for state standards, and are both developed and reviewed by trained educators throughout the process.
How is the performance assessment scored?	Performance tasks are scored by local teachers (i.e., teachers at the students' own school) using: (1) detailed rubric guidance that is specific to each question/task being scored and (2) model responses to each question, intended to support teacher understanding of less common student responses to open-ended tasks. Samples of teacher-scored work are submitted to the state for auditing and validation purposes.
What knobs does this assessment prioritize?	Disciplinary sophistication, transfer.

New York State Regents Exams

Regents: Social Studies Essay (sample materials)

ENDURING ISSUES ESSAY

This question is based on the accompanying documents. The question is designed to test your ability to work with historical documents. Some of these documents have been edited for the purposes of this question. As you analyze the documents, take into account the source of each document and any point of view that may be presented in the document. Keep in mind that the language and images used in a document may reflect the historical context of the time in which it was created.

Directions: Read and analyze each of the five documents and write a well-organized essay that includes an introduction, several paragraphs, and a conclusion. Support your response with relevant facts, examples, and details based on your knowledge of social studies and evidence from the documents.

An enduring issue is a challenge or problem that has been debated or discussed across time. An enduring issue is one that many societies have attempted to address with varying degrees of success.

Task:

- Identify ***and*** explain an enduring issue raised by this set of documents
- Argue why the issue you selected is significant ***and*** how it has endured across time

In your essay, be sure to

- Identify the enduring issue based on a historically accurate interpretation of *at least* ***three*** documents
- Explain the issue using relevant evidence from *at least* ***three*** documents
- Argue that this is a significant issue that has endured by showing:
  - How the issue has affected people or has been affected by people
  - How the issue has continued to be an issue or has changed over time
- Include relevant outside information from your knowledge of social studies

In developing your answer to Part III, be sure to keep these explanations in mind:

Identify—means to put a name to or to name.

Explain—means to make plain or understandable; to give reasons or causes of; to show the logical development or relationship of something.

Argue—means to provide a series of statements that provide evidence and reasons to support a conclusion.

New York State Regents Exams

Document 1

> . . . The spinning of cotton into threads for weaving into cloth had traditionally taken place in the homes of textile workers. In 1769, however, Richard Arkwright patented his 'water frame', that allowed large-scale spinning to take place on just a single machine. This was followed shortly afterwards by James Hargreaves' 'spinning jenny', which further revolutionised the process of cotton spinning.
>
> The weaving process was similarly improved by advances in technology. Edmund Cartwright's power loom, developed in the 1780s, allowed for the mass production of the cheap and light cloth that was desirable both in Britain and around the Empire. Steam technology would produce yet more change. Constant power was now available to drive the dazzling array of industrial machinery in textiles and other industries, which were installed up and down the country. . . .

Source: Matthew White, "The Industrial Revolution," The British Library online, October 14, 2009

Document 2

> . . . The first important economic changes in Japanese life came in communication and transportation. In both these fields the changes helped national defense as well as economic growth.
>
> The first field to be improved was communications, which is easier to develop than transportation. In 1869 a telegraph line ran from the lighthouse outside Yokohama harbor to the center of the city. Thus the keeper could send early news of ships' arrivals to the people who had to meet them. By 1872 telegraph lines linked Tokyo to Sapporo in the north and to Nagasaki on the west. From Nagasaki undersea cables to Shanghai and Vladivostok tied Japan to the rest of the world. In fact, the telegraph helped the imperial government crush the Satsuma rebellion in 1877. It also helped businessmen in their jobs. However, most people seldom used the telegraph, and then only for very special or urgent messages.
>
> A new postal system grew out of the old Tokugawa system of runners between major cities. In 1871 relays of runners ran the 343 miles (about 556 kilometers) between Tokyo and Osaka in 78 hours. The next year Tokyo had mail delivery three times a day. In 1874 Japan joined the Universal Postal Union, a link to other major countries. Soon, inexpensive postcards became a favorite way of sending personal messages. By the end of Meiji, post offices handled about 33 pieces of mail each year for each Japanese individual. Quick communication had become a way of life. . . .

Source: John F. Howes, *Tradition in Transition: The Modernization of Japan*, Macmillan Publishing Co., 1975

Document 3

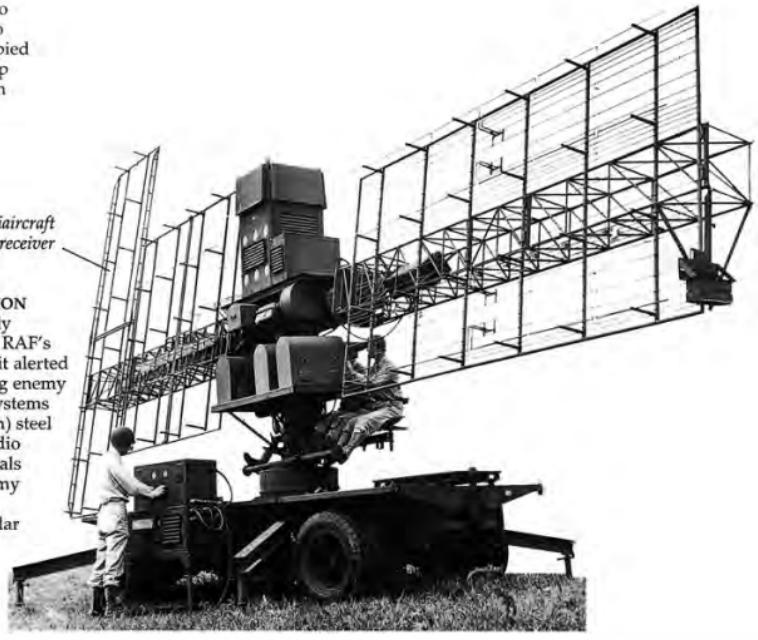
World War II

Homemade wireless receiver used by a Dutch family during the occupation

SECRET RADIO
Hidden in a tin, this radio was used by a Dutch family to listen to the BBC (British Broadcasting Corporation). Broadcasts included general war news, messages from the exiled Dutch royals, and coded messages to secret agents. Owning a radio was forbidden in many occupied countries, but that did not stop people from making their own and using them in secret.

US Army field telephone

PORTABLE COMMUNICATIONS
Field telephones were used by Allied and Japanese soldiers to keep in contact with their commanders and the rest of their unit. The speed of the Japanese advance across Southeast Asia and the Pacific meant that troops needed efficient communications to inform headquarters of their progress, and on the whereabouts of the enemy.



ENEMY DETECTION
Radar was hugely important to the RAF's success because it alerted them to incoming enemy aircraft. Radar systems used 300-ft (90-m) steel masts to emit radio signals. The signals bounced off enemy planes and were picked up by radar receivers. This alerted pilots to scramble into the air to resume battle.

Document 4

The Golden Quadrilateral is a 3,633-mile expressway designed to link India's four largest cities. The project was launched in 2001.

> . . . Announced in 1998 by then Prime Minister Atal B. Vajpayee, who is credited with giving the project its grandiose [impressive] name, the Golden Quadrilateral is exceeded in scale only by the national railway system built by the British in the 1850s. For decades after its 1947 independence, India practiced a kind of South Asian socialism in keeping with the idealism of its founders, Gandhi and Nehru, and its economy eventually stalled. In the 1990s the country began opening its markets to foreign investment, led by a pro-growth government and staffed by an army of young go-getters who speak excellent English and work for a fraction of the wages paid in the West. Yet India's leaders realized their decrepit [decaying] highways could hobble the country in its race toward modernization. "Our roads don't have a few potholes," Prime Minister Vajpayee complained to aides in the mid-1990s. "Our potholes have a few roads.". . .
>
> It's not unusual, on a six-lane superhighway, to find oxcarts, water buffalo, motorcycles, and the occasional line of trucks and cars coming straight at you, in your lane, driving the wrong way because it's shorter or easier or perhaps because they're confused. Goats graze the median strip, and traffic is often held up by sacred cows, the only users of the highway that seem oblivious to the danger flying around like shrapnel. . . .

Source: Don Belt, "Fast Lane to the Future," *National Geographic*, October 2008

Document 5

The Middle East: Worries about privacy—a concern of the elite

> . . . In many Middle Eastern countries, most people can't afford a computer, let alone access to the Internet, so the security and privacy of information is generally not a concern. Some, however, who have access to this technology—primarily citizens of oil-rich countries—are perturbed [upset] about the security of their personal information when using the Internet. Social networks such as *Facebook* and *MySpace* are growing in popularity with young people in the region, although they are banned in some countries, such as Syria. Privacy, however, is tenuous [questionable]; it is generally a risky proposition to challenge unauthorized governmental intrusion into one's computer.
>
> Although the Egyptian interior ministry in July 2002 formed a specialized unit to combat Internet-related crimes, offenders are primarily punished according to the country's laws governing communications and the protection of intellectual property. That is the situation in most Middle Eastern countries.
>
> There is very limited use of biometric security* in Middle Eastern countries, although some airports, including Israel's Ben-Gurion Airport, have for years used iris-scanning biometrics to verify the identities of employees and passengers. . . .

Source: Larry Greenemeier, "International Report: What Impact Is Technology Having on Privacy around the World?," *Scientific American* online, August 18, 2008

New York State Regents Exams

OPTIONAL PLANNING PAGE

Enduring Issues Essay Planning Page

You may use the Planning Page organizer to plan your response if you wish, but do NOT write your essay response on this page. Writing on this Planning Page will **NOT** count toward your final score.

My Enduring Issue is:_____

Essay Requirements	Yes	Circle documents that apply	One or two possible ideas for outside information
Is this an issue supported by *at least **three*** documents? Which documents support this issue?		1 2 3 4 5	
Which documents can be used to develop the explanation for this issue?		1 2 3 4 5	
Has this issue significantly affected people or been affected by people? In which document or documents do you see this?		1 2 3 4 5	
Has this issue endured across time or changed over time? In which document or documents do you see this?		1 2 3 4 5	

Source: Sample tasks and information about scoring and use for all disciplines can be found at the New York State Education Department, Office of State Assessment website at www.nysedregents.org (accessed 09/06/2024).

Table A5. Smarter Balanced Assessment Consortium Performance Tasks

Smarter Balanced Assessment Consortium Performance Tasks	
What does the performance assessment look like?	Performance tasks are included in both math and ELA. These tasks are designed to focus on the integration of content and skills across standards; require students to engage in planning, interaction, and interpretation; and generate a product to address an authentic scenario or problem. In math, performance tasks look like using mathematical concepts and practices to address real-world problems, such as making a recommendation for product design based on the amount and cost of materials or determining how quickly different candles will burn. In ELA and literacy, performance tasks involve text analysis and the development of an argumentative, explanatory, narrative, opinion, or informational essay. Students are given a scenario that motivates "finding out more" and sources to read and interpret, and are then asked to engage in a purposeful writing task.
How is the performance assessment used?	Performance tasks are administered in the classroom during the second semester of the year. Each performance task includes a short classroom activity to help all students access the task, followed by a short individual performance task. The performance task contributes score points to students' final summative score, in conjunction with items delivered through an on-demand computer-adaptive test.
How is the performance assessment developed?	The performance assessments are developed in collaboration between testing experts and teachers, using extensive item specifications and task formats designed to provide evidence for SBAC's assessment claims while meeting criteria for high-quality performance tasks within the SBAC system. These criteria include qualitative and quantitative features of stimuli (e.g., scenario features, text complexity), expectations for scaffolding and on-ramping to the task, developmentally appropriate range of performance elicited, specificity of task and scoring to the targets being elicited, and language and visual design considerations to ensure accessibility. For complete performance task specifications, see these documents provided by SBAC.

Smarter Balanced Assessment Consortium Performance Tasks	
How is the performance assessment scored?	Tasks are centrally scored by expert reviewers, including educators. Each task is scored against a detailed rubric designed for the task type and claims being assessed. Rubrics include the learning objectives assessed, scoring criteria, and narrative description for making a determination about score points awarded. SBAC performance task scoring also leverages anchor responses, which are example responses at each level of performance. These responses have been initially screened by expert reviewers trained by SBAC on using the rubrics and are further validated by educators across a range of stakes and contexts.
What knobs does this assessment prioritize?	Disciplinary sophistication, Transfer

Source: Smarter Balanced Assessment Consortium. Sample tasks can be found at https://sampleitems.smarterbalanced.org/ and information about scoring and use can be found at https://validity.smarterbalanced.org/ (accessed 09/06/2024).

Table A6. Washington Civics Assessment

Washington Civics Assessment	
What does the performance assessment look like?	These tasks ask students to consider an authentic and developmentally appropriate civics issue and develop written and oral presentations in response to the task. While tasks vary in terms of topics and exact skills assessed, each asks students to consider multiple ideas and perspectives, and to use evidence-based reasoning to support their final product. Students are given a rubric as part of the task to support understanding expectations and their own monitoring of their learning.
How is the performance assessment used?	The performance assessments are primarily intended to support teaching and learning. Teachers can use the tasks how and when they think them most appropriate to classroom instruction, and task implementation support includes additional resources (e.g., diagrams, readings, discussion suggestions) to facilitate situating the task in an authentic learning experience. The tasks and rubrics are designed to model and support civics instruction in the classroom.
How is the performance assessment developed?	Performance tasks are developed through a collaboration between the state education agency and educators in the Washington State social studies cadre. Tasks are designed to meet state standards in social studies, informed by national frameworks, as well as key standards in other related content areas (e.g., ELA).
How is the performance assessment scored?	Performance tasks are scored by local teachers for use in their classrooms. Schools and districts are required to verify that such tasks are being used at least once per grade-band (4–5th, 7–8th, and 11–12th grades). To do so, districts submit a verification report to the state.
What knobs does this assessment prioritize?	Disciplinary authenticity, disciplinary sophistication, integration across disciplines.

Assessment Task

Teacher's Instructions to Students

1. Say: "Today you will take the Washington OSPI-developed social studies assessment for economics (Grades 6–8). This assessment is called *Meeting Needs and Wants*."
2. Provide the class with copies of the student's section of the assessment (which may include the student's task, response sheets, rubrics, templates, and glossary), along with any other required materials.
3. Tell the students that they may highlight and write on these materials during the assessment.
4. Have the students read the directions to themselves as you read them aloud. We also encourage you to review the glossary and scoring rubric with the students.
5. Answer any clarifying questions the students may have before you instruct them to begin.

Accommodations

Refer to the student's IEP or 504 plan.

Student's Copy of the Task

The following section contains these materials for students:

- the student's task: *Meeting Needs and Wants* (Grades 6–8)
- assessment rubric
- worksheets and handouts (optional)

Washington Civics Assessment

Meeting Needs and Wants

As a citizen and member of a community, you need to have an understanding of different economic systems. You will analyze the methods that societies use to meet the economic needs and wants of citizens.

Your Task

In a cohesive paper or presentation, you will:

- ☐ State a claim regarding how two groups met their needs and wants. You should include an analysis of how the two groups' approaches are similar or different.
- ☐ Present the background on your claim, including:
  - An explanation of how the rights, values, and/or beliefs of both groups have affected their economic choices.
  - Two or more specific examples (at least one example for each group).
- ☐ Provide reason(s) for your position, including an analysis of one or more significant similarities or differences related to the groups' ways of meeting their needs and wants.
- ☐ Refer explicitly in the paper or presentation to three or more credible sources that provide relevant information:
  - Cite your sources when you draw information from them: for instance, when you summarize, paraphrase, or quote, and when you refer to facts, figures, and ideas.
  - Provide complete publication information for each source in your bibliography or list of works cited.

Washington Civics Assessment

Meeting Needs and Wants Rubric

Score	4	3	2	1
Claim/Position	I stated a claim regarding how two groups met their needs and wants; I included: An analysis of how the two groups' approaches are similar or different. A general statement of how these groups' experiences help us understand a current issue or event.	I stated a claim regarding how two groups met their needs and wants and included an analysis of how the two groups' approaches are similar or different.	I stated a claim regarding how two groups met their needs and wants, but I did not analyze how the two groups' approaches are similar or different.	I stated a claim that is vague, implausible, or inaccurate.
Concepts	I provided background on the claim by: Explaining how rights, values, and/or beliefs of both groups have affected their economic choices: I included two examples (at least one for each group). AND Utilizing the concept of opportunity cost to explain how the groups met their needs and wants.	I provided background on the claim by explaining how rights, values, and/or beliefs of both groups have affected their economic choices: I included two examples (at least one for each group).	I provided background on the claim by explaining how the groups' rights, values, and/or beliefs have affected their economic choices: I included one example.	I described the ways the groups met their needs and wants, but I included no specific examples.
Evidence/Reasoning	I provided reason(s) for the claim, supported by evidence; my reasoning includes: An analysis of significant similarities and/or differences related to their ways of meeting their needs and wants. Three or more examples.	I provided reason(s) for the claim, supported by evidence; my reasoning includes: An analysis of significant similarities and/or differences related to their ways of meeting their needs and wants.	I provided reason(s) for the claim, supported by evidence; my reasoning includes: An analysis of significant similarities and/or differences related to their ways of meeting their needs and wants. One example.	I provided reason(s) for the claim, but included no supporting evidence. OR I provided an inaccurate analysis of significant similarities and/or differences related to their ways of meeting their

Washington Civics Assessment				
Sources	I referred explicitly in the paper or presentation to four or more credible sources that provide relevant information.	I referred explicitly in the paper or presentation to three credible sources that provide relevant information.	I referred explicitly in the paper or presentation to two credible sources that provide relevant information.	I referred explicitly in the paper or presentation to one credible source that provides relevant information.
ELA Citations	I properly cited sources in the paper, presentation, and bibliography, using a specific format.	I adequately cited sources in the paper, presentation, and bibliography.	I minimally cited sources in the paper, presentation, and bibliography.	I incorrectly cited sources in the paper, presentation, and bibliography.

Source: Washington Office of Superintendent of Public Instruction. *OSPI-developed social studies assessments* (accessed 09/06/2024).

Appendix B: Frequently Asked Questions

1. **What are performance assessments?**

   They are assessments that reveal what students really know and can do through rigorously observing students' performance of complex and meaningful tasks.

   All assessments try to surface evidence of student thinking and understanding. In performance-based assessments, students are asked to demonstrate their ability to use the targeted knowledge and skills as authentically and directly as possible. The performance serves as direct evidence of the degree to which students have grasped the targeted knowledge and skill. Performance-based approaches can be used for all kinds of assessment purposes, including formative processes that support immediate learning, intervening checks on individual and group progress, end-of-learning evaluation, and demonstration of students' abilities. High-quality performance assessments include clear success criteria and rubrics for evaluating student thinking.

2. **Why should I use performance assessments?**

   Because they create better—more relevant, more usable—information and offer better preparation for students' postsecondary lives.

   Reasons that system leaders should consider using performance assessments include:

   - **Better Data About Student Understanding and Abilities.** The goal of this assessment approach is to more closely emulate the real-world conditions within which students have to use the knowledge and abilities they develop in school. In doing so, performance assessments surface information about the most important kinds of student thinking so that students, families, and educators can make more informed decisions about next steps.

   - **Meaningful Experiences for All.** When systems include performance assessments, they ensure that every single student taking the assessment has the opportunity to engage in the kinds of rich learning experiences performance assessments entail (e.g., engaging in research, evaluating evidence to construct an argument, conducting investigations to make sense of observations). In the short term, this means that every student gets at least a handful of these experiences. In the long term, these tasks encourage educators to shift their day-to-day teaching and learning to better prepare students for richer assessments, improving educational experiences for every learner. Performance assessments can be extremely motivating, and systems that include them in authentic and meaningful ways often see more diverse students participating and succeeding in high-level coursework.

   - **Robust and Coherent Professional Learning.** When teachers are involved in developing and scoring performance tasks, they have an opportunity to better understand their disciplines, how students learn and perform, and how to support a wide range of students in taking next steps.

3. **Do all performance assessments look the same?**

   No, but they do have some common features.

- High-quality performance assessments vary in format, but they are all designed to meet specific (and specified) performance targets; include rubrics and success criteria for self, peer, and teacher evaluation of student work; and surface a range of student thinking through direct performance on open-ended tasks that ask students to generate original artifacts.
- Performance assessments vary in format for many reasons:
  - Like all assessments, performance assessments must be designed for particular purposes. Different purposes need different formats (e.g., consider an assessment designed to check whether students understood a specific idea versus an assessment designed to surface student ability to creatively apply ideas from an entire course).
  - Because high-quality performance assessments are authentic applications of ideas and skills as they would be used in the real world, performance-based approaches will necessarily vary based on what is being targeted. A performance assessment addressing an understanding of properties of matter might take the form of a citizen science project coordinated across the community while a performance assessment addressing students' grasp of literary analysis skills might be an original critique of a newly published work.

4. **Do performance assessments serve a particular purpose? (Or, how should I use them?)**

   Performance assessments can be designed to serve a wide range of purposes, because they represent an approach to assessment generally, rather than a specific type of assessment.

   Performance-based approaches to assessment surface how students are thinking, rather than whether they know specific facts or steps of a procedure. As a result, performance-based approaches can be useful in formative learning and growth processes by guiding both teachers' and students' decisions through real-time, actionable information, as well as serving as summative measures of student learning. Performance-based assessments are designed such that students are learning by engaging with the assessment, rather than the assessment being separate from learning.

5. **How do performance assessments support all learners?**

   Performance assessments provide opportunities to address multiple equity lenses. This includes limiting bias and sensitivity, supporting multilingual learners and students with special needs, and offering opportunities to attend to student identity and agency.

   High-quality performance assessments are considered more equitable because they can exhibit many features of equity that standardized measures cannot easily attend to. These features include:

   - **More Accurate and Fuller Understanding of Student Thinking.** Performance assessments allow students to demonstrate how they are thinking, rather than simply giving a right (or wrong) answer. This can enable students who may have a strong conceptual understanding of a target to make what they *do* know and what they *can* do visible, allowing better feedback and support for learning and growth. Because performance assessments are not overly reliant on vocabulary or isolated pieces of knowledge, they can be more supportive for learners from a variety of linguistic

and cultural backgrounds, and at a variety of points along learning progressions. Specifically, performance assessments can be supportive of multilingual learners and students with special needs because they can be appropriately scaffolded and supported, provide opportunities to use home language, and demonstrate understanding through multiple modalities.

- **Attending to Multiple Cultural Ways of Knowing.** Performance-based approaches can allow students to make their thinking visible in ways that are connected to their home lives and the identities they bring to the classroom.
- **Building Confidence and Engagement in the Assessment Through Choice and Decision-Making.** Performance assessments can provide students with opportunities to exercise choice and decision-making, from more limited opportunities like choosing which examples or data to cite, to more expansive degrees of choice, like what projects to take on, specific topics to explore, or how to demonstrate their thinking. In all cases, the opportunity to make their own choices can be empowering, engaging, and confidence-boosting for all learners.
- **Increasing Relevance to Student Lives.** Well-designed performance-based assessments are relevant to students' lives—they are grounded in real-life scenarios that are meaningful to the learners engaged in the assessment and authentic examples of how and why the targeted knowledge and skills are used in the real world; engaging in the assessment provides an experience of value for students.

  Bottom line: By giving students opportunities to demonstrate what they *do* know and what they *can* do through supportive assessments that don't rely on catching students in the act of not knowing something, assessments are more likely to surface more accurate and complete information about student understanding and abilities than are multiple choice questions students might get right or wrong for a host of unrelated reasons. Importantly, *all* students, including lower-performing students, can engage with performance assessments; well-designed performance tasks include appropriate scaffolds and supports such that all students can show what they know, and teachers have precise information about how to support every student's progress.

6. **Are performance assessments as valid as the multiple choice tests I'm used to?**

   In fact, high-quality performance assessments often yield scores that are more valid than those of standardized assessments.

   When we talk about "validity" in assessments, we are trying to capture how confident we can be in whether an assessment score represents the knowledge, skills, and abilities we want to learn about from student performance. Multiple choice questions, like those used on most standardized assessments, rely on students choosing the right answer from a list of possibilities; whether students get them right or wrong can depend heavily on vocabulary skills, test preparation skills, cueing from incorrect answer choices, etc. At worst, these features that determine success have nothing to do with what we're trying to measure; at best, they are superficial proxies for deeper understanding that multiple choice measures alone can't surface. High-quality performance-based

approaches, in contrast, give students the opportunity to demonstrate the knowledge, skills, and abilities we're interested in learning about directly and deeply, offering a much more comprehensive and trustworthy window into student progress.

7. **Will test scores increase if I incorporate performance assessments into the summative test?**

   Maybe, maybe not—but they will surface trustworthy data that can support better decisions moving forward in either case.

   Performance assessments measure student learning in different ways, making it a little complicated to determine an exact relationship to student performance on traditional multiple choice tests.

   - **Higher Scores.** In some instances, students may achieve higher scores on performance-based assessments than on traditional multiple choice tests. Research suggests that for these students, performance assessments do reveal student brilliance that was masked by traditional tests because they: (1) allow students to more completely show their conceptual understanding, (2) motivate students to engage and persevere with the task, and (3) often provide opportunities for students to reflect and iterate on their work, providing opportunities for growth within the assessment itself.

   - **Lower Scores.** In other instances, students may earn lower scores on performance assessments. Performance tasks are less easy to "game" with test-taking strategies, or rapid review and cram sessions that motivate memorization for a short period of time. They also unmask student thinking; for example, students who may be able to identify mitochondria as the powerhouse of the cell on a multiple choice test may not be able to predict what would happen to a person's health if they developed a mitochondrial disease. For some students, this may mean that performance assessments deflate artificially high scores on traditional assessments by revealing the edge of what students know and can do. This may result in lower scores, but it also provides better data about student thinking.

   - **No Change in Scores.** For many students, performance on rich tasks will be correlated with their performance on traditional test measures. What is gained through performance tasks are rich opportunities and signals for teaching and learning.

8. **If performance assessments are open-ended, how can they be used in comparable, reliable ways?**

   For those assessments that require high degrees of comparability, open-ended tasks can be used very effectively.

   Strategies for effectively using open-ended tasks:

   - **Clearly Defined Common Rubrics.** By creating thoughtful, consistently applied success criteria and rubrics for student responses, open-ended tasks can be evaluated reliably across students and by educators, focusing on those features that are being evaluated while providing students with the flexibility to respond in a range of ways that demonstrate those features.

- **Common Assessment Tasks.** Some assessment situations might require going beyond common rubrics to common tasks themselves. These tasks can be designed to be open-ended while ensuring that all learners engaged in a particular task make their thinking visible in similar formats and contexts.

- **Calibrated Scoring Procedures.** In systems that use performance assessments, there is substantial investment in calibration among raters. This includes extensive training (and sometimes certification) of anyone involved in scoring student work, routine calibration opportunities, back-end audits of local scoring, and validation of scoring through other measures.

- It is important to remember that not all assessments need to be able to provide comparisons among students. In many instances, it might be far more useful (to teachers, students, and parents) to understand how students are progressing along a learning progression rather than how they compare to one another.

9. **If I want to use performance assessments, do I have to entirely abandon other kinds of assessment?**

   Of course not! While performance assessments can be hugely valuable when used as a primary approach to and philosophy for assessment, students should experience a variety of assessments and feedback opportunities throughout their K–12 learning experiences.

   This might include occasional multiple choice questions/tests that supplement richer performance-based approaches; equity-oriented surveys of student experience that monitor student interest, identity, and agency considerations directly; and both formal and informal performance-based approaches that themselves take a variety of forms. When making decisions about which assessment types to use, remember to consider:

   - **Purpose of Assessment.** What kinds of information about student thinking will be most useful for my current purpose, and how can I surface that information most effectively? What do I want students, educators, and other stakeholders to be able to do with the assessment?

   - **Signaling.** What are my assessment decisions signaling to students and parents about what is most valuable about student learning and performance?

   - **Story Over Time.** While individual assessments can tell us a lot about student thinking, they should be part of a body of work over time that tells a story of student learning and progress. Where does any given assessment fit into this story, and how can we design and implement it such that it is surfacing and providing the best possible information at that time, for that purpose?

Endnotes

1. Marion, S. F., Pellegrino, J. W., & Berman, A. I. (2024). *Reimagining balanced assessment systems*. The National Academies Press; Ruiz-Primo, M., Li, M., Wills, K., Giamellaro, M., Lan, M., Mason, H., & Sands, D. (2012). Developing and evaluating instructionally sensitive assessments in science. *Journal of Research in Science Teaching*, *49*(6), 691–712. http://dx.doi.org/10.1002/tea.21030; National Research Council. (2014). *Developing assessments for the Next Generation Science Standards*. The National Academies Press. https://doi.org/10.17226/18409. It should be noted that this guidance is in direct opposition to the stated intent of the Every Student Succeeds Act and associated peer review guidance, which require state assessments to produce individual student scores and diagnostic information that supports the learning and growth of each student.

2. Badrinarayan, A. (2024). *Design principles for instructionally relevant assessment systems.* Learning Policy Institute. https://doi.org/10.54300/111.528

3. Bennett, R. E., & Gitomer, D. H. (2009). Transforming K–12 assessment: Integrating accountability testing, formative assessment and professional support. In C. Wyatt-Smith & J. J. Cumming (Eds.), *Educational assessment in the 21st century* (pp. 43–61) Springer; Black, P., & William, D. (2007). Large scale assessment systems: Design principles drawn from international comparisons. *Measurement: Interdisciplinary Research and Perspectives*, *5*(1), 1–53; Lane, S. (2009). Issues in the design and scoring of performance assessments that assess complex thinking skills. In G. Schraw & D. Robinson (Eds.), *Assessment of higher order thinking skills* (pp. 265–302). Information Age Publishing; Messick, S. (1996). Validity of performance assessments. In G. W. Phillips (Ed.), *Technical issues in large-scale performance assessment* (pp. 1–18). National Center for Education Statistics; Shavelson, R. J., Davey, T., Holland, P. W., Webb, N. M., & Wise, L. L. (2015). *Psychometric considerations for the next generation of performance assessment*. Educational Testing Service.

4. Guha, R., Wagner, T., Darling-Hammond, L., Taylor, T., & Curtis, D. (2018). *The promise of performance assessments: Innovations in high school learning and college admission*. Learning Policy Institute.

5. Kane, M., Crooks, T., & Cohen, A. (1999). Validating measures of performance. *Educational Measurement: Issues and Practice*, *18*(2), 5–17.

6. Wei, R. C., Pecheone, R. L., & Wilczak, K. L. (2014). *Performance assessment 2.0: Lessons from large-scale policy and practice*. Stanford Center for Assessment, Learning, and Equity.

7. Darling-Hammond, L. (2017). *Developing and measuring higher order skills: Models for state performance assessment systems*. Learning Policy Institute and Council of Chief State School Officers; Darling-Hammond, L., Ancess, J., & Falk, B. (1995). *Authentic assessment in action: Studies of school and students at work*. Teachers College Press.

8. Darling-Hammond, L. (1994). Performance-based assessment and educational equity. *Harvard Educational Review*, *64*(1), 5–30. https://doi.org/10.17763/haer.64.1.j57n353226536276

9. Federal peer review guidance (critical element 6.4) requires that assessments provide information that can be used to address the specific needs of students. Efforts to meet this peer review criterion often involve efforts to produce reports and subscores from large-scale tests that are ultimately not well positioned to provide instructionally useful information, simply because the test was not designed to provide educators with the information they need in the first place. Performance assessments are often better positioned to provide this kind of information, and they can be an innovative and more useful path to meeting this criterion when used as part of state summative assessments. It should be noted that in order for performance assessments to maximize their instructional impact, it is important that educators see more than an aggregate score that reflects student proficiency—they need access to the tasks, and at least some student work, to be able to make meaning of student thinking processes.

10. Darling-Hammond, L., & Adamson, F. (2010). *Beyond basic skills: The role of performance assessments in achieving 21st century standards*. Stanford Center for Opportunity Policy in Education; Lane, S. (2010). *Performance assessment: The state of the art*. Stanford Center for Opportunity Policy in Education.

11. National Urban League and UnidosUS. (2022). *Education assessment, accountability, and equity: 2022 Phase 1 final report of the Future of Assessment and Accountability Project*. https://nul.org/sites/default/files/2023-04/FOAA_Final%20Phase%201%20Report_April2023_final%20.pdf

12. Systematic and well-supported use might include dedicated and sustained professional learning, access to tasks and surrounding curriculum-implementation support, instructional coaches who can help teachers reflect on and improve practice with performance tasks, and infrastructure within schools, districts, and states that supports collaborative learning among teachers.

13. Darling-Hammond, L., & Adamson, F. (2010). *Beyond basic skills: The role of performance assessments in achieving 21st century standards*. Stanford Center for Opportunity Policy in Education; Darling-Hammond, L., Pecheone, R., Jacquith, A., Schultz, S., Walker, L., & Wei, R. C. (2010). *Developing an internationally comparable balanced assessment system that supports high-quality learning*. National Conference on Next Generation Assessment Systems. https://www.ets.org/Media/Research/pdf/Darling-HammondPechoneSystemModel.pdf; Pecheone, R. L., & Kahl, S. (2010). *Through the looking glass: Lessons learned and future directions for performance assessment*. Stanford Center for Opportunity Policy in Education.

14. Note that this paper is focused on the use of authentic performance tasks in conjunction with more standardized assessment items, as most systems do.

15. Lane, S. (2010). *Performance assessment: The state of the art*. Stanford Center for Opportunity Policy in Education.

16. Mislevy, R. J. (1996). Test theory reconceived. *Journal of Educational Measurement*, *33*(4), 379–416; Mislevy, R. J., & Haertel, G. D. (2006). Implications of evidence-centered design for educational testing. *Educational Measurement: Issues and Practice*, *25*(4), 6–20.; Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2003). On the structure of educational assessments. *Measurement: Interdisciplinary Research and Perspectives*, *1*(1), 3–62.

17. Fine, M., & Pryiomka, K. (2020). *Assessing college readiness through authentic student work: How the City University of New York and the New York Performance Standards Consortium are collaborating toward equity.* Learning Policy Institute. https://learningpolicyinstitute.org/product/assessing-college-readiness-authentic-student-work-report

18. Krajcik, J., Schneider, B., Miller, E., Chen, I.-C., Bradford, L., Bartz, K., Baker, Q., Palincsar, A., Peek-Brown, D., & Codere, S. (2021). *Project-based learning increases science achievement in elementary school and advances social and emotional learning.* George Lucas Educational Foundation. https://www.lucasedresearch.org/wp-content/uploads/2021/01/MLPBL-Research-Brief-1.pdf

19. Saunders, M., Kaul, M., & Thompson, C. (2020). *Using performance assessments to support student learning in Los Angeles Unified School District.* Learning Policy Institute. https://learningpolicyinstitute.org/sites/default/files/product-files/CPAC_Performance_Assessments_Student_Learning_CS_LAUSD_REPORT.pdf

20. Darling-Hammond, L., & Falk, B. (2014). Supporting teacher learning through performance assessment. In L. Darling-Hammond & F. Adamson (Eds.), *Beyond the bubble test: How performance assessments support 21st century learning* (pp. 207–236). Jossey-Bass.

About the Author

Aneesha Badrinarayan was the Director of State Performance Assessment Initiatives at the Learning Policy Institute (LPI), where she supported states in developing meaningful assessment systems as a force for better teaching and learning. Badrinarayan is a national expert in innovative assessment systems, with a focus on meaningful science assessments. Her portfolio includes authoring more than 100 resources designed to support educators and leaders in the development of better assessment systems; leading the development of the 2028 Science Framework for the National Assessment of Educational Progress; leading collaborative national efforts to redefine "alignment" in the era of new state standards; partnering with states to develop instructionally relevant assessment systems; and leading multiple multistate collaboratives focused on better assessments in service of student learning. Badrinarayan also served as the Director of Special Initiatives at Achieve, a museum educator, and a bench scientist. She earned an MS in Neuroscience at the University of Michigan, where she served as a research fellow for the National Institute of Mental Health, and earned a BA in Biology from Cornell University.

1530 Page Mill Road, Suite 250
Palo Alto, CA 94304
p: 650.332.9797

1100 17th Street, NW, Suite 200
Washington, DC 20036
p: 202.830.0079

learningpolicyinstitute.org

The Learning Policy Institute conducts and communicates independent, high-quality research to improve education policy and practice. Working with policymakers, researchers, educators, community groups, and others, the Institute seeks to advance evidence-based policies that support empowering and equitable learning for each and every child. Nonprofit and nonpartisan, the Institute connects policymakers and stakeholders at the local, state, and federal levels with the evidence, ideas, and actions needed to strengthen the education system from preschool through college and career readiness.