Design Principles for Instructionally Relevant Assessment Systems

Aneesha Badrinarayan

LEARNING POLICY INSTITUTE

OCTOBER 2024

Acknowledgments

The author is grateful to the many policymakers, education leaders, researchers, and educators who contributed to our thinking by sharing their experiences with past and current assessments, as well as their hopes for the future, with us. In particular, I thank Erin Furtak, Christopher Harris, Joe Krajcik, Scott Marion, James Pellegrino, Bill Penuel, and Jill Wertheim for their time and technical expertise in guiding the questions and providing ongoing feedback throughout the development of these principles. Although all content, views, and recommendations made here are those of the Learning Policy Institute (LPI) alone, the following people helped make this report a reality by sharing their experiences, successes, and challenges with efforts to create more learning-centered assessment systems: Daniel Alcazar-Roman, Horatio Blackman, Matt Blomstedt, Julianna Charles Brown, David Cook, Sara Cooper, Nathan Dadey, Elena Diaz-Billelo, Paul Dumas, Ellen Ebert, Cory Epler, Carla Evans, Denise Forte, Jeff Greig, Jeremy Heneger, Ellen Hume-Howard, Alissa Kilpatrick, Angela Landrum, James Lane, Paul Leather, Bob Lenz, Tana Luther, Susan Lyons, Rachael Manzer, Melissa Mendenhall, Nicholas Munyan-Penney, Dawn Novak, Michael Novak, Lillian Pace, Susan Patrick, Raymond Pecheone, William Penuel, Stephen Pruitt, Jennifer Randall, Brian Reiser, Breigh Rhodes, Sam Ribnick, Jeffrey Riley, David Ruff, Lorrie Shepard, Hal Smith, Michele Snyder, Corrine Steever, Rhonda True, Katie Van Horne, Audrey Webb, Justin Wells, Gene Wilhoit, Ted Willard, and members of the State Performance Assessment Learning Community and the Interstate Learning Community, who so graciously shared their insights and perspectives.

I also thank my LPI colleagues Linda Darling-Hammond, Michael DiNapoli, Tara Kini, Tiffany Miller, Patrick Shields, Charlie Thompson, and Julie Woods for their thought partnership around issues of assessment policy and practice. I am indebted to the members of the LPI Communications team for their invaluable support in designing, producing, 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 authors and not those of our funders.

External Reviewers

This report benefited from the insights and expertise of three external reviewers: Scott Marion, Executive Director at the Center for Assessment; Jim Pellegrino, Professor at the University of Illinois Chicago; and Jill Wertheim, Director at SCALE@WestEd. I thank them for the care and attention they gave the report.

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

This report can be found online at https://learningpolicyinstitute.org/product/design-principles-assessment.

This work is licensed under the Creative Commons Attribution-NonCommercial 4.0 International License. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc/4.0/.

Document last revised October 11, 2024

Table of Contents

Executive Summary iv

Introduction 1

Lessons From the Field 5

Design Principles for Instructionally Relevant Assessment Systems 9

Defining State Assessment Systems 9

Rationale and Critical Features 12

Authentic 12

Curriculum-Anchored 15

Educative 19

Developmental and Asset-Oriented 22

Reflective of and Responsive to Learners 25

Useful for Informing Decisions That Impact Instruction 28

Systems in Progress: Assessment Models and Examples 31

Systems That Are Reimagining What—and How—to Assess in Large-Scale Assessment Systems 33

Systems That Are Centering Teacher Growth and Learning via Assessment Systems 40

Systems That Are Changing How Students See Themselves and Make Their Thinking Visible in Large-Scale Assessments 41

System Conditions for Instructionally Relevant Assessment Systems 44

Conclusion 46

Endnotes 47

About the Author 49

List of Figures and Tables

Figure 1 Ways to Anchor Assessment to Curriculum 16

Figure 2 Snapshot of a Sample Summative Assessment Task 34

Figure 3 Snapshot of a Sample Classroom Performance Task 35

Figure 4 AP Computer Science Principles Classroom Performance Task 37

Figure 5 Sample Writing Task Leveraging Hot and Warm Reads 39

Table 1 Educator Insights About Instructional Impact of State Assessments 1

Table 2 Instructionally Relevant Decisions That Interest Holders Make Based Partially on Assessments 4

Table 3 Design Principles for Instructionally Relevant Assessment Systems 10
Table 4 Characteristics of Assessment Systems That Center Authenticity 13
Table 5 Characteristics of Assessment Systems Anchored to Curriculum 17
Table 6 Characteristics of Assessment Systems That Are Educative 20
Table 7 Characteristics of Assessment Systems Designed to Attend to Student Assets and Development Over Time 23
Table 8 Characteristics of Assessment Systems Designed to Be Reflective of and Responsive to Learners 26
Table 9 Characteristics of Assessments That Are Useful for Informing Decisions That Impact Instruction 29
Table 10 Examples Connecting Intended Action With Assessment System Design 31

Executive Summary

The decisions states make regarding what their assessments look like and what kind of information they produce inevitably shape instruction. Since the No Child Left Behind Act of 2001 ushered in an era of testing-based accountability for schools, state assessments have been governed by a set of design decisions that emphasize easily generated, easily compared scores—even when these assessments are somewhat superficial proxies for the rich performance expectations state standards set for student learning. This makes sense if state assessments play a narrow and siloed role, focused on sending up a red flag around school performance and triggering a cascade of follow-up actions. While this might be consistent with how designers *intend* for assessments to be used, there have been unfortunate and unintended consequences for teaching and learning.

Many states want to construct assessment systems more deliberately to achieve their goals. By leveraging our understanding of how various groups use information from state assessments, we can design assessments and systems that have a net positive impact on instruction. We can do so by designing assessments that:

- Define what shifts from current instructional practices should be incentivized.
- Ensure that state assessments are designed and communicated such that the most proximate logical way to match the state assessment in local practice (e.g., interim assessments, classroom assessment resources) mirrors activities that reflect research on how students develop disciplinary knowledge and practice.
- Recognize that what happens in the classroom is not limited to interactions between teachers, students, and the content of instruction alone. What and how students learn is shaped by decisions made by educators and leaders throughout the system. Indeed, many of the instructionally relevant decisions that state assessments are most likely to influence lie outside day-to-day teacher–student interactions. By focusing on the most impactful ways state assessments influence instruction, states can ensure that assessments have a positive influence on instruction without extending into purposes that large-scale external assessments are not well suited to address.
- Provide teachers and leaders with information that offers a significant perceived value-add over other kinds of information they already receive through their classroom, school, and district instructional and assessment practices and resources.

Design Principles for Instructionally Relevant Assessment Systems

Based on evidence from assessment system design and implementation as well as lessons learned working alongside various states, a set of design principles emerge that govern assessments intended to support teaching and learning. These principles are designed to:

- Build upon current conceptions of alignment to standards.
- Focus on the most discerning features of assessment system design—that is, those features that are most likely to distinguish between systems that lead to positive shifts in instruction and those that have neutral or negative impact on teaching and learning, while allowing for a range of ways states could enact these principles.

- Triangulate among the most important instructional shifts; the key users; and the specific, evidence-based behaviors we want to influence.
- Walk the line between aspirational and achievable—it is unlikely that any state's current large-scale assessment program meets all of these design principles, but it is imminently conceivable that they could make different design decisions right now to bring their assessments into better alignment with instructionally impactful goals.

Instructionally relevant assessment systems are intentionally designed to incorporate the following six principles:

1. **Authentic.** Assessments should highlight and center the key concepts, modes of inquiry, and ways of learning in the discipline.
2. **Curriculum-Anchored.** Assessments are designed such that high-quality curriculum better prepares students for success on the assessment, the assessment incentivizes the adoption and use of high-quality curriculum, and the assessment supports implementation of high-quality materials.
3. **Educative.** Assessments build educator and student understanding of and experience with high-quality teaching and learning in the discipline.
4. **Developmental and Asset-Oriented.** Assessments recognize what students *do* know and *can* do, and surface progress relative to individual student performance and along appropriate learning progressions.
5. **Reflective of and Responsive to Learners.** Assessments follow principles of universal design and cultural responsiveness to ensure that each learner is supported in making their thinking visible.
6. **Useful for Informing Decisions That Impact Instruction.** Assessments are designed to produce relevant information at appropriate times to support decision-making.

By centering features of assessments that support better student learning experiences, teacher practice, and systematic supports and decision-making, we can create assessment systems that have a net positive impact on instruction. The design principles detailed in this report reflect ambitious but accomplishable goals for assessment systems—and large-scale systems, including states as well as national and international programs, are already on the path to making this work a reality. As systems move forward, keeping "positive instructional impact" as the North Star and centering decisions on specific instructional shifts from the current state of teaching and learning that assessments should support can help system designers make the best decisions within their local contexts.

Introduction

The decisions states make regarding what their assessments look like and what kind of information they produce inevitably shape instruction. Since the No Child Left Behind Act of 2001 ushered in an era of testing-based accountability for schools, state assessments have been governed by a set of design decisions that emphasize easily generated, easily compared scores—even when these assessments are somewhat superficial proxies for the rich performance expectations state standards set for student learning. This makes sense if state assessments play a narrow and siloed role, focused on sending up a red flag around school performance and triggering a cascade of follow-up actions. While this might be consistent with how designers *intend* for assessments to be used, there have been unfortunate and unintended consequences for teaching and learning.

For example, in a recent conversation about how state assessment systems—defined as all instruments, supports, and communications from the state education agency about assessment—influence teaching and learning, teachers, principals, and instructional coaches reported a number of impacts. (See Table 1.)

Table 1. Educator Insights About Instructional Impact of State Assessments

Impact on instruction	Examples from educators
Making decisions about instructional scope and sequence, including which units within instructional materials to exclude from instruction, based on the expected [a] blueprint of the state assessment	• "I cannot finish my full [high-quality instructional materials] within a given school year, so I decide which units to do a lighter touch on or skip altogether based on what will be covered on the test." • "I always try to make sure we do at least one project that is meaningful to students, but I usually try to do that after the testing window is over."
Deciding the kinds of practices used more frequently in classroom activities [b]	• "We focus on reading passages and exit tickets that use prompts like 'find the main idea' rather than reading whole books, because that's what is on our [interim and state] tests." • "Our state test asks students to fill in boxes for scientific modeling, so I modify modeling activities in my curriculum to give students practice with that." • "We prioritize individual work because student discourse is not assessed on the state test."

Impact on instruction	Examples from educators
Choosing instructional content and contexts	• "Our assessment focuses on sort of generic, everyday phenomena, like balls rolling down ramps or boiling water, so those are the phenomena I prioritize in my science classroom." • "I want to choose [texts] that are more relevant to my specific students, but I want to make sure they are exposed to the kinds of texts and content that will be on the test, so I usually prioritize that."
Making decisions about local assessment resources	• "We chose an interim assessment system that seemed best positioned to give students and teachers practice with the summative and was most likely to be predictive of student scores on the summative." • "I encouraged my district to purchase [specific formative assessment item bank] because most of the questions are released questions from our state test."

[a] Either because of explicit state communication or because of perceived trade-offs educators expected to be reflected on the state assessment.

[b] Many educators also noted that this impact, in particular, was especially relevant for newer or more novice teachers. Given recent experiences with teacher turnover and influxes of new educators into classrooms, leaders report that they are paying increased attention to this impact.

Source: Learning Policy Institute. (2024).

Educators are clear: While test designers and assessment theorists may focus on narrow, well-defined roles for different kinds of assessment, the reality is that (1) all interest holders in the system look to the state assessment as an indicator of what should be happening in schools and classrooms, and (2) local decision-makers often try to make local instructional contexts match the signals sent by state assessments as closely as possible.

These impacts on instruction are logical, should be expected, and are in some cases deeply concerning—but current assessments are not designed to even consider this kind of instructional impact. The unintended consequence is that state assessments often have a negative impact on teaching and learning, narrowing curriculum decisions and encouraging teachers and leaders to make decisions that seem more likely to maximize scores, even if they are inconsistent with evidence related to how to best support student learning and what learning to support. These impacts are further reinforced each time a school or district purchases interim assessments that match the summative assessment or chooses to spend professional learning days unpacking state assessment items or writing assessment tasks that mirror current state assessment items.

Many states want to construct state assessment systems more deliberately to achieve their goals. By leveraging our understanding of how various groups use information from state assessments, we can design assessments and systems that have a net positive impact on instruction. We can do so by designing assessments (see Table 2) that:

> By leveraging our understanding of how various groups use information from state assessments, we can design assessments and systems that have a net positive impact on instruction.

- Define what shifts from current instructional practices should be incentivized.
- Ensure that state assessments are designed and communicated such that the most proximate logical way to match the state assessment in local practice (e.g., interim assessments, classroom assessment resources) mirrors activities that reflect research on how students develop disciplinary knowledge and practice.
- Recognize that what happens in the classroom is not limited to interactions between teachers, students, and the content of instruction alone. What and how students learn is shaped by decisions made by educators and leaders throughout the system (Table 1). Indeed, many of the instructionally relevant decisions that state assessments are most likely to influence lie outside day-to-day teacher–student interactions. By focusing on the most impactful ways state assessments influence instruction, states can ensure that assessments have a positive influence on instruction without extending into purposes that large-scale external assessments are not well suited to address.
- Provide teachers and leaders with information that offers a significant perceived value-add over other kinds of information they already receive through their classroom, school, and district instructional and assessment practices and resources.

In many conversations about the potential role and impact of large-scale assessments, meaningful discussions about instructional impact are held hostage by narrowly defined notions of instructional utility that suggest that assessments can only have an impact on instruction if they can be used by teachers to change in-the-moment instruction. It is almost as though two different conversations are being had at the same time: On the one hand, teachers and leaders are bemoaning the consequences many current large-scale assessment designs have on instruction, including sacrificing depth of conceptual learning for a broad survey of facts and privileging passive recitation of memorized information over active engagement in disciplinary practice and knowledge building. On the other hand, assessment experts sometimes reject the call for instructionally sensitive assessments by arguing that large-scale or external assessments cannot help teachers decide how to change their lesson plans in a given week—even though that is not what teachers, leaders, or students are identifying as the problem large-scale assessments pose for instructional experiences.

Recognizing this tension, the design principles developed and discussed here focus on a broader conceptualization of instructional impact of state assessment systems. Put simply, this report is concerned with how states can establish assessment systems that help educators, students, and leaders make a range of decisions that have a net positive impact on teaching and learning experiences and outcomes.

Table 2. Instructionally Relevant Decisions That Interest Holders Make Based Partially on Assessments

Interest holder	Examples of instructionally relevant decisions and actions
Students	• Reflecting on learning goals and progress (metacognition) • Codesigning learning experiences that are relevant and meaningful • Actively engaging in disciplinary inquiry • Revising work to meet shared goals and expectations • Providing feedback on learning experiences
Teachers	• Providing opportunities for deep, sustained, and compelling learning • Enacting culturally and linguistically responsive teaching and reteaching practices that deepen disciplinary skills and understanding while supporting the development of disciplinary identities • Leveraging students' current understanding and experience as a foundation within which to anchor new learning • Engaging students in learning that mirrors the authentic behaviors and conceptual development of the discipline • Learning more about student learning
School and district leadership	• Scheduling decisions that enrich curricular opportunities for all learners (e.g., inclusive learning for emerging multilingual learners rather than pulling them out of class time for language remediation, providing sustained and coherent time for science and social studies in K–5) • Using observation protocols and educator coaching that reflect the major instructional shifts of the discipline • Implementing equitable grading policies that match how learning toward standards should occur (e.g., allowing grading frequency and approach to be consistent with how students develop and make progress visible in high-quality teaching and learning, focusing standards-based grading approaches on deeper learning targets rather than superficial coverage) • Adopting and implementing high-quality curriculum and instructional materials • Investing in systemwide curriculum-based professional learning • Establishing responsive course options and pathways that create opportunities for learners
State leadership	• Developing and adopting coherent instructional materials, assessments, and professional learning policies • Incentivizing the use of high-quality instructional materials (e.g., reimbursement for the purchase of these materials, support for professional learning for high-quality open educational resources) • Tailoring state-offered professional learning to support key instructional shifts

Source: Learning Policy Institute. (2024).

Lessons From the Field

Research on assessment design and implementation, along with a series of conversations with state and local leaders, assessment and instructional materials designers, educators at the classroom and regional levels, and research and implementation partners, illuminates how large-scale assessments often function relative to instruction. These insights suggest that some lessons about assessments can be positioned to be instructionally supportive. Lessons include:

- **Focus on assessment as part of teaching and learning systems.** For more than 20 years, state assessments have been seen first and foremost as a partner to current accountability systems, providing the lion's share of input into formulas used to rate schools and allocate resources and interventions accordingly. This orientation has led to certain trade-offs in state assessment design (e.g., sacrificing depth of student understanding for breadth of content coverage, emphasizing large numbers of easily scored items vs. richer open-ended tasks). States seeking to design assessment systems that center instruction must make a different set of trade-offs. Rather than asking "How can we get the most efficient data on student proficiency?" they should want to consider organizing their assessment systems around questions like "How can we produce data on student achievement that can be used to drive meaningful, positive change in decisions about curriculum and professional learning?"

- **Consider how test content and formats influence the nature of instruction.** Leaders we talked with in nearly every state say that their standards, and the assessments aligned to them, represent the minimum of what all students should know and be able to do. In teaching and learning, educators and leaders are encouraged to take steps to help students soar beyond the benchmarks for student progress established by state policies. However, teachers share that the reality is that in most cases, instruction shifts to meet and reflect, not exceed, what is represented on assessments. This becomes problematic when assessments—often perceived as the operationalization of state standards—are superficial proxies for much richer expectations. Such assessments functionally lower the bar that we want all students to meet. This often has the greatest impact on the students most in need of support, undermining the goal of high standards for all students.

- **Prioritize the most important instructional shifts.** State assessments send powerful signals and are one of the few ways that states can influence instruction in every school and classroom within their borders. At the same time, state assessments cannot do everything, and they are not a replacement for the comprehensive information that teachers get from daily instruction. Recognizing which shifts from current practice are the most essential for more meaningful student learning can help states design assessments that are particularly impactful on teaching and learning.

  For example, several studies found that when states started including writing prompts and essay writing on their English language arts assessments, teachers began focusing on writing in the classroom.[1] In science, teachers and partners shared that while there has been some success in shifting toward making sense of phenomena and problems with science ideas, there is still an overemphasis on knowing or restating science facts and ideas due to the perceived focus of state assessments on knowing disciplinary ideas. Essential shifts from current practice that teachers, leaders, and science education researchers identified for science instruction during the development of these design principles included: (1) more authentic engagement in science and engineering

practices and crosscutting concepts in service of rich sensemaking, (2) making sense of relevant phenomena and complex problems that require components of multiple performance expectations to address and may not have a clear answer, and (3) leveraging culturally responsive teaching practices to ensure that all students have opportunities to excel in science. Accomplishing these shifts would require a different emphasis in science assessment, shifting away from recitation of science ideas and toward their use in service of meaning-making about real-world phenomena and problems.

- **Use assessments that inform a range of decisions and actions that are relevant to instruction, from many different actors.** While improving the utility of state assessments for teachers is essential, state assessments are used by a wide range of interest holders throughout the system to make decisions that influence instruction (see Table 2). Indeed, while teachers and partners often bemoan the impact of state assessment on teaching and learning, they share that two reasons are driving this impact:

  1. Teachers feel obligated to prepare students for the test, even if the test is understood to be inconsistent with deeper learning.

  2. School or district administrators—not teachers—are making decisions that shape instruction in well-meaning but misguided efforts to boost scores, such as adopting instructional materials that match standards but ignore important pedagogical supports, or requiring incoherent, one-off professional learning workshops.

  In many cases, assessment experts respond to these insights from teachers by emphasizing that state summative assessments are not *intended* to shape instruction and should not be used to do so—but the reality is that, best practice or not, the field repeatedly tells us this *is* how tests are used. Being transparent about who state assessments are actually used by, and what actions they take based on assessment results, can help states develop assessments that produce the information most likely to lead to decisions that enable and support better teaching and learning.

- **Consider instructional materials in conjunction with assessments.** Traditional approaches to state assessment are somewhat agnostic to curriculum decisions, operating from the assumption that proficiency data will be sufficient to support schools' and districts' programmatic decision-making. However, research suggests that state assessments would be both more valid and more useful if they were intentionally coherent with high-quality curriculum. Evidence from the learning sciences demonstrates that what and how students learn often directly influences how they make their thinking visible on an assessment.[2] Moreover, research also suggests that access to high-quality instructional materials—materials that operationalize the growing body of work on how students learn—has had a significant impact on improving teacher practice and student outcomes.[3] Taken together, there is a compelling argument for state assessments that are more intentionally connected to high-quality curriculum, such that (1) the assessment results are more trustworthy, and (2) they support and incentivize high-quality teaching and learning through the principled implementation of high-quality instructional models, materials, and approaches.

There is a compelling argument for state assessments that are more intentionally connected to high-quality curriculum.

This inherently implies that states will have to pay more attention to—and play a more significant role in supporting—curriculum decisions that happen locally. A growing number of states are already playing more explicit roles in curriculum design and instructional materials selection by providing guidance, financial incentives, and requirements for the adoption and implementation of high-quality instructional materials.[4] These emerging examples highlight a different kind of role for state education agencies to play in instructional materials selection and emphasize the importance of reciprocal and supportive partnerships between state and local education agencies. As Breigh Rhodes of the Louisiana Department of Education described it:

> For too long, [state education agencies] have had a singular kind of role: Control and direct. Anything that we don't tell districts or schools, or teachers what to do, we just have ignored or said was outside of our purview. That doesn't have to be our story. ... We need to reimagine our roles into one of partnership and support. We should be paying attention to curriculum decisions, not to tell districts what to do, but to understand what's going on, why those decisions are being made, and what they need—so *we* can change our actions to be supportive of *them*. We can work *together* to navigate decisions about instructional materials so that our students and teachers get access to the best possible supports.

- **Make sure assessments are experienced as a significant enough value-add to teachers, students, and families that they overcome the burden of their implementation.** Implementation of external assessments requires effort on the part of teachers, schools, and districts. As states seek to provide educators and leaders with better assessments, it is essential that teachers, students, and families see these assessments as a value-add—that the benefits of engaging in the assessment itself as well as the information yielded significantly outweigh the burden of implementation. Striking this balance is particularly important as states consider more realistic assessment elements, such as curriculum-embedded performance tasks. State assessments need to be consistent enough with what teachers are already doing that they are valid measures of student learning while also providing enough new information about student thinking to warrant spending the time and effort to implement the test.

  The specific benefits that are particularly compelling may differ by context, depending on the current state of instruction (e.g., access to high-quality instructional materials, strength of the teaching workforce, existing professional learning structures, historical and cultural elements of education in a state or region). In our conversations with educators as part of developing these design principles, they identify some common elements as offering a significant value-add over their current assessment systems, including:

  - transparent access to high-quality tasks that are engaging and relevant to learners, exemplify practice for teachers, and are a substantive component of state assessments;
  - access to student work and educative rubrics that highlight how to make meaning of student thinking and support progress;
  - tools and resources for teaching and learning, connected to outcomes;

- classroom- and school-level reporting that focuses on instructional shifts and strategies;
- access to interest, experience, and other opportunities to learn data alongside assessment results; and
- professional learning opportunities for making sense of rich student information (e.g., scoring student work on performance tasks).

- **Harness the potential of generative artificial intelligence (genAI) for reimagining assessment systems, not reifying problematic practices.** The availability of relatively sophisticated artificial intelligence tools is changing the landscape of trustworthy measures of student performance in ways that deeply impact assessment—in terms of both what student learning we surface and the way we do it. For example, new advancements in genAI and other related technologies may make it easier to implement more authentic and sophisticated assessments at scale. Current efforts to use genAI to support scoring and meaning-making around student work, to facilitate coaching and professional learning for educators, and to be embedded within more immersive tasks themselves may position states to support much more meaningful assessment systems in more efficient ways in the near future.

  Importantly, as genAI is poised to offer significant innovations to assessment development and scoring and reporting, it will be essential that innovators use instructional relevance and equity considerations as first-order principles when deciding whether and how to incorporate genAI into assessment processes. GenAI should be used to make assessment systems more instructionally relevant, not to reinforce practices that have the potential to cause harm to students and their learning. For example, instead of using large-language models to generate large numbers of superficial multiple-choice items or to help teachers and schools create more selected-response assessments, developers could consider using the same technology to help customize deeper learning tasks to local contexts and operationalize asset-based scoring and reporting of performance assessments at scale and on more immediate timelines.

Design Principles for Instructionally Relevant Assessment Systems

Based on evidence from assessment system design and implementation as well as lessons learned by the Learning Policy Institute and members of the State Performance Assessment Learning Community while working within, alongside, and across states, a set of design principles emerge that govern assessments intended to support teaching and learning. (See Table 3.) These principles are designed to:

> Based on evidence and lessons from experience, a set of design principles emerge that govern assessments intended to support teaching and learning.

- Build upon current conceptions of alignment to standards.
- Focus on the most discerning features of assessment system design—that is, those features that are most likely to distinguish between systems that lead to positive shifts in instruction vs. those that have neutral or negative impact on teaching and learning, while allowing for a range of ways states could enact these principles.
- Triangulate among the most important instructional shifts, the key users, and the specific, evidence-based behaviors we want to influence.
- Walk the line between aspirational and achievable—it is unlikely that any state's current large-scale assessment program meets all of these design principles, but it is imminently conceivable that they could make different design decisions right now to bring their assessments into better alignment with instructionally impactful goals.

The following sections provide an overview of the design principles, followed by more detail on each principle, including rationale for the design principle and critical features that states may consider including in their assessment system design to enact the principle within their systems. Because many of the states pursuing innovative, instructionally relevant assessment systems currently are doing so in science, many of the design principles that follow include specific examples from science contexts. However, these design principles can and should govern any disciplinary assessment seeking to prioritize instruction first and foremost.

Defining State Assessment Systems

This report uses the frame of "state assessment systems" to refer to the collection of assessment instruments, supporting documents, and assessment-related services that states directly provide to districts, schools, and teachers with an expectation that their use will influence or change student performance on summative measures of achievement or proficiency used by the state. In some states, this might be concentrated on a single instrument, such as the end-of-year assessment, or it may occur in a through-year model. In other states, this might encompass a combination of an on-demand statewide test with locally selected instruments (e.g., performance assessments). The principles are intentionally

focused on features of state summative assessments rather than optional local activities because practitioners and decision-makers have stressed that *without* important changes to the state's formal assessment system—the system that educators and local leaders feel accountable to and measured by—local efforts cannot be as impactful.

Importantly, while these design principles have been developed to be ambitious but achievable for states, they are strong indicators of the potential for positive instructional impact for many different kinds of large-scale assessment systems, such as those used in districts, in school networks, and as part of national systems (e.g., Advanced Placement, International Baccalaureate). (See Table 3.)

Table 3. Design Principles for Instructionally Relevant Assessment Systems

Design Principle	Summary statement
Authentic. Assessments should highlight and center the key concepts, modes of inquiry, and ways of learning in the discipline.	The assessment system should include authentic tasks that represent ambitious examples of learning and performance in the discipline. These tasks should reflect sophisticated and complete performances, signal and support engagement with science and engineering practices (SEPs) and crosscutting concepts (CCCs), and center sensemaking around meaningful phenomena and problems. Importantly, these tasks should engage students in sensemaking in ways that are expected in science and reflect the most important instructional shifts we want to see. This may include both individual and collaborative work; cascades and bundles of SEPs and CCCs; student choice, either among tasks or about how to engage within a task; and more.
Curriculum-Anchored. Assessments are connected to, and informed by, high-quality curriculum.	Assessments should signal, incentivize, and support the use of high-quality curriculum that center active engagement with the disciplines in ways that operationalize evidence from the learning sciences about how disciplinary knowledge and practice are developed. This positions assessments to provide information that can be particularly useful to instruction; encourage the use of instructional materials and models that focus on deeper learning; and provide students with assessments that are, in and of themselves, meaningful learning experiences. In some cases, assessments may be designed to closely reflect high-quality instructional materials (HQIM); in other cases, assessments may be designed to be coherent with HQIM but focus on complementing existing curriculum (e.g., providing extended transfer opportunities, providing opportunities to better attend to broader issues).

Design Principle	Summary statement
Educative. Assessments build educator understanding of effective teaching and how students learn in the discipline.	Assessments—the tasks, student data, and supports for interpretation—should build educator understanding of what high-quality disciplinary teaching and learning look like, what kinds of tasks can develop and evaluate that learning, and how to provide feedback in ways that support progress toward these goals (e.g., state standards). Importantly, assessments attend carefully to the learning of teachers and students alike and are designed such that teachers too feel like they have learned something meaningful about their practice through the implementation and examination of assessments. This means that what assessments signal, measure, and provide information about should directly speak to the kinds of actions and decisions we want students, educators, and leaders to make—and help them learn both how to do so and why it is important. This may be accomplished by incorporating performance tasks into the instructional process; releasing items, tasks, and student work so that educators can see the kinds of tasks students are being asked to accomplish and what scores reflect; involving educators in designing and scoring tasks; providing task and student response annotations; providing concrete next steps to take, aligned to features of high-quality teaching and learning in science and based on student performance profiles; and making student experience data available to educators and leaders to contextualize performance.
Developmental and Asset-Oriented. Assessments recognize what students *do* know and *can* do and surface progress relative to students' own performance and along appropriate learning progressions.	Assessments should focus on providing all students with opportunities to show what they know and can do relative to sophisticated disciplinary meaning-making. This includes emphasizing scoring and reporting that focuses on recognizing facets of student understanding and supports student growth over time. Assessments should also provide information about student performance along extended, multiyear learning progressions as well as expected learning progressions within a learning sequence (e.g., accounting for how modeling is expected to develop).

Design Principle	Summary statement
Reflective of and Responsive to Learners. Assessments follow principles of universal design and cultural responsiveness to ensure that each learner is supported in making their thinking visible.	Assessments should reflect students' cultural and linguistic experiences, employ multiple modalities for acquiring information and working through tasks, and include opportunities for students to demonstrate their learning in a variety of ways.
Useful for Informing Decisions That Impact Instruction. Assessments are designed to produce relevant information at appropriate times to support decision-making.	Assessment data must be made available at times when it can be used to positively impact instruction. In some cases, this might look like getting assessment data to users in a more timely fashion, particularly if the assessment design is intended to support changes in instruction for the specified cohort of students. However, it should be noted that timing is not necessarily a discerning feature—more timely assessment results are useful to instruction only if the information is designed to be supportive of instructional decisions at those intervals (e.g., through-year assessment design). In other cases, assessments may be designed primarily to help teachers reflect on their own practice and plan for improving their instruction for their next cohort of students. In these cases, states may intentionally decide to slow down the process of returning scores to students to allow teachers to engage in rich and educative scoring experiences that can have a direct impact on instruction but result in scoring on a slower cadence than more automated processes might produce.

Source: Learning Policy Institute. (2024).

Rationale and Critical Features

Authentic

Assessments should highlight and center the key concepts, modes of inquiry, and ways of learning in the discipline.

Why is this important?

State assessment efforts can serve at least three important functions in teaching and learning systems, including:

1. Signaling what students should know and be able to do as a result of instruction, aligned to the state's standards, portraits of a graduate, and other visioning policies and documents.
2. Providing an example of the kinds of experiences students should be engaging in as part of ambitious teaching and learning practices.

3. Providing information about how students are progressing toward expectations in ways that students, families, teachers, and leaders can use to influence instructional practices in positive and productive ways.

All three functions require that students, families, teachers, and leaders can: (1) trust that state assessments are measuring the knowledge, skills, and abilities that are most important to the discipline; (2) look to state-provided assessments as exemplars of what student performance should look like; and (3) easily use the information and examples provided by the state assessment to backward map to what teaching and learning should look like. For example, in science, this means that if we want district, school, and classroom practice to cultivate authentic sensemaking and problem-solving with science and engineering practices, crosscutting concepts, and disciplinary ideas, we need state assessment systems to include authentic opportunities for students to do so. Similarly, if we want students to make meaning of texts and write to think and synthesize in English classes, we need assessments that provide students with opportunities to do so at the level of sophistication we expect in the classroom, incorporating processes of drafting and revision that aim for deep analysis. Without state assessments that are worth teaching to—that highlight and center the most important disciplinary activities as authentically as possible—many students will simply never experience high-quality instruction, as the educators and leaders responsible for their learning get caught in never-ending cycles of trying to game the state test. (See Table 4.)

> Without state assessments that are worth teaching to, many students will simply never experience high-quality instruction.

Table 4. Characteristics of Assessment Systems That Center Authenticity

Assessment systems that center authenticity are less likely to ...	Assessment systems that center authenticity are more likely to ...
• Use only selected-response item types that prioritize contrived activities or behaviors as a proxy for performance (e.g., selecting from a list when the targeted performance is synthesizing information). • Provide superficial or highly contrived contexts that are disconnected or unnecessary to the standards, competency, or performance being assessed. • Focus on assessing isolated and decontextualized knowledge and skills. • Prioritize breadth and equal coverage of every standard.	• Demonstrate developmentally appropriate versions of what disciplinary professionals do. • Use tasks that represent authentic uses of the disciplinary activities as they would be applied in the real world. • Assess sensible bundles of standards that work together to contribute to a substantive and impactful performance. • Prioritize depth and sufficient coverage such that the assessment accurately represents the intent of the standards.

Source: Learning Policy Institute. (2024).

Critical Features of This Design Principle

1. **Use high-quality performance tasks.** Assessments should leverage high-quality performance tasks as a component of state assessment systems. These tasks should focus intentionally on aspects of standards that are essential to students' disciplinary learning but are difficult to assess on traditional assessments that rely heavily on selected-response items. For example, these aspects may include:

    - centering tasks on robust phenomena, problems, and scenarios that ask students to apply multiple disciplinary practices, crosscutting concepts, and conceptual ideas (potentially across disciplines);
    - collaborative investigations, modeling, and argumentation;
    - student-driven research, synthesis, and communication of findings; and
    - student thinking in response to phenomena and problems that contain authentic uncertainty.

2. **Emphasize sensemaking and meaning-making over trying to cover the breadth of standards.** Supporting student sensemaking[5] and meaning-making is at the core of disciplinary standards in math, science, English language arts, and social studies and should be centered in state assessment design. In science, assessments should center sensemaking as the primary construct for individual items, blueprints, and performance descriptors, focusing on eliciting evidence of the degree to which students can reason with the targeted ideas and practices. This is in contrast to assessment designs that attempt to isolate evidence of the breadth of each dimension. (See Reflective of and Responsive to Learners for more information about the influence of this principle on measuring students across performance levels.) In science, for example, this may mean that assessments lean heavily into realistic engagement with the science and engineering practices and crosscutting concepts to help support the shift away from rote content memorization and toward meaningful sensemaking.

3. **Focus on the most-needed shifts and higher-leverage aspects of standards appropriate to the grade band.** State assessments are a relatively blunt instrument to guide change, and well-meaning efforts to do it all have resulted in assessments that are relatively unsuccessful in compelling changes to instructional practices. Unfortunately, this goes beyond ineffectiveness, often encouraging changes that actually make deeper, sustained learning less likely—this includes encouraging instruction that prioritizes superficial and quick coverage of disparate content instead of deeper focus on the most important content and practices of the discipline. States should carefully consider the highest-priority shifts state assessments can support and ensure that those same elements are prioritized in the assessment design. For example, one of the most widely acknowledged shifts needed in science teaching and learning across grade levels is more support and deeper student engagement in science and engineering practices, particularly around interpreting and using evidence. Accordingly, states may design assessments that foreground science and engineering practices, emphasizing using scientific ideas as part of making sense of provided data and distinguishing between the validity of different claims and sources. In this example, states may choose to emphasize these elements over other types of performance, such as questions that focus on comprehensively surfacing conceptual understanding of a given science idea. Across math, English language arts,

> and science, another major shift researchers and practitioners emphasize is the need for students to integrate knowledge and skills across many different standards. Assessments could be designed to signal this kind of shift by developing items and tasks that emphasize measuring the construct that emerges at the intersection of multiple standards, even if this means defining new performance expectations for assessment that are derived from state standards statements.
>
> Similarly, states should consider which elements of the standards are going to (1) have the biggest impact on instruction, and (2) be most critical for future learning when prioritizing how standards are translated to assessments. The most important shifts and elements of standards may be slightly different across grade bands. State assessments should balance the needs of particular grade bands with vertical coherence across K–12.

Curriculum-Anchored

Assessments are connected to, and informed by high-quality curriculum.

Why is this important?

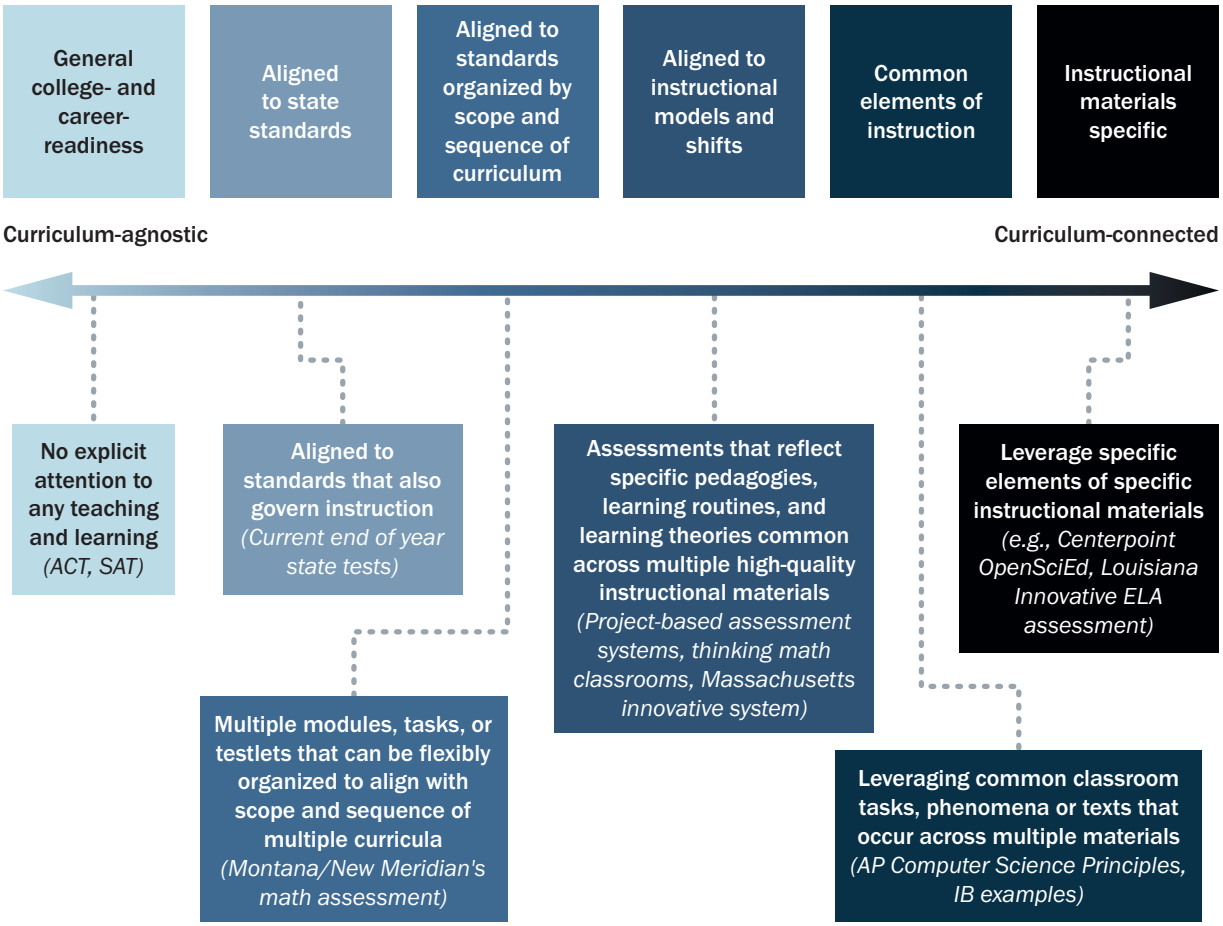
As one state leader stated:

> If we want assessments to be more instructionally relevant, we have to pay attention to what's going on in highly effective classrooms and make sure our assessments encourage those activities and behaviors.

Traditional approaches to assessment design assume a unidirectional flow of influence in which state standards drive curriculum and assessment design independently, and, by virtue of alignment to standards, the assessments and curriculum are aligned. This "transitive property of alignment" approach makes some sense in theory, but it is not borne out by practice. Evidence from the learning sciences tells us that how students learn matters, and learning experiences influence how students make their thinking visible.[6] As a result, "curriculum agnostic"[7] assessments that focus on end-of-instruction performance aligned to the letter of the standards, without attending to how students developed those ideas and practices, often result in superficial assessments that are not particularly useful to any high-quality curriculum context.

Instead, states should consider designing assessments that intentionally leverage features of high-quality curriculum.[8] Across disciplines, leaders and developers have taken great care to ensure that high-quality instructional materials[9] (HQIM) and associated professional learning reflect the best evidence we have on how students learn within each of the science disciplines and do so in service of building mastery toward state science standards. There are a range of approaches—some more tightly coupled to specific instructional materials and others more loosely coupled to curriculum via common features of HQIM like theories of learning and instructional models—that states could explore to design "curriculum-anchored assessments" in service of better teaching and learning across their state systems (see Figure 1).

Figure 1. Ways to Anchor Assessment to Curriculum



Sources: Adapted from Badrinarayan, A., & Steiner, D. (2023). *Positioning state assessment systems in service to teaching and learning: The role of high-quality curriculum in state assessment design*. Education First; Dadey, N. & Badrinarayan, A. (2022, April). In search of the "just right" connection between curriculum and assessment: Considering options between curriculum-specific and curriculum-agnostic state assessment [Blog post]. *Center for Assessment*.

By considering curriculum and assessment together, states can provide assessments that are more easily used by educators and leaders and ultimately have a stronger and more positive influence on curriculum implementation for students. Done well, these curriculum-anchored state assessments can measure knowledge, skills, and abilities in ways that attend to what and how students have had the opportunity to learn in science. These same assessments can also (1) incentivize the use of high-quality curriculum approaches and materials by making clear that these approaches will help students be successful on state assessments, and (2) provide robust and ongoing support for connecting student progress to instructional decisions—ultimately acting as a tool to support local implementation of high-quality curriculum in ways that support every learner across diverse contexts.

It should be noted that this design principle assumes that state assessments should be more sensitive to instruction—in other words, performance on assessments provided by the state should vary based on whether students have received high-quality, impactful instruction, with students who have had

better instruction performing better on state assessments, and students who have had more limited opportunity to learn performing less well. It has been well documented that increasingly external, large-scale assessments are generally less sensitive—and may be *in*sensitive—to instruction.[10] While we do not take issue with this finding, this design principle calls for us to challenge this moving forward—state assessments *should* reflect to some degree whether students have had opportunity to experience and reap the benefits of high-quality teaching and learning. (See Table 5.)

Table 5. Characteristics of Assessment Systems Anchored to Curriculum

Assessment systems that are anchored to curriculum are less likely to ...	Assessment systems that are anchored to curriculum are more likely to ...
• Emphasize decontextualized facts or skills. • Be intentionally agnostic to the range of students' lived experiences. • Be disconnected from meaningful teaching and learning practices—intentionally instructionally insensitive. • Report student proficiency in ways that are disconnected from learning and assume mastery rather than learning.	• Center and emphasize the major instructional, content, and performance shifts of the discipline (e.g., emphasize figuring out phenomena as the mechanism of learning and assessment, using disciplinary practices as part of meaning-making, focusing on content-rich reading comprehension approaches). • Reflect the kinds of teaching and learning experiences that students should be experiencing in the curriculum. • Connect meaningfully with what students have had opportunity to learn—both in terms of contextualized assessment and interpretations of transfer and generalizability. • Leverage common instructional materials as a window into shared classroom experiences that can be leveraged as part of assessment design.

Source: Learning Policy Institute. (2024).

Critical Features of This Design Principle

1. **Anchor to curriculum in ways that reflect intended and likely use.** There are many ways to make meaningful connections between high-quality curriculum and a state's assessment system. The most appropriate connection will depend on the specific way a state intends for the assessment to influence instruction (e.g., incentivizing the uptake of HQIM or supporting the development of particular instructional practices connected to desired shifts) as well as historical and cultural contextual factors within the state (e.g., a history of strong state-provided support for instructional materials selection vs. a culture of teacher-led local curriculum development). States should consider both intended and likely uses to tailor the curriculum connection to maximize positive instructional shifts.

When considering how best to connect with curriculum to support state goals, states should consider how the assessment design will be both consistent with and complementary to high-quality curriculum. Striking the right balance between consistent and complementary features is essential to making sure the assessment has the desired impact on instruction. Features that are consistent with the curriculum should help educators and leaders clearly see the connections between the assessment and high-quality curriculum and should support easy translation between information surfaced by the assessment and changes to instructional and programmatic practices and decisions. Features that complement existing high-quality curriculum should provide a significant value-add to teachers beyond what they already have access to in using HQIM.

2. **Provide explicit guidance for how the assessment approach can be used with multiple high-quality instructional materials available within the state.** No state mandates curriculum, nor is it likely that any single curriculum will be implemented in the same way over time. Therefore, states should be clear about how their assessment system is compatible with multiple curricula while maintaining a strong stance on features of high-quality curriculum that the assessment signals and measures. State actions might include:

    - clearly communicating assumptions about learning and the underlying theory of learning governing assessment system design;
    - using curriculum surveys to identify the top three to five HQIM being used within a state and providing insight into the kinds of inferences that can be made about student performance relative to expected learning experiences;
    - building a library of assessment tasks that assess the same targeted standards but are intentionally designed for different HQIM and instructional models (e.g., project-based learning models, 5E instructional models, storyline curriculum models) for teachers to use based on their local curriculum decisions;[11] and
    - coupling state-provided tasks with clear navigation routines that support educators making transitions from curricular contexts into and out of the assessment, to support coherence from a student perspective.

3. **Offer curriculum-anchored assessment professional learning.** One of the most powerful aspects of a curriculum-anchored assessment system is that it provides a clear avenue for coherent curriculum, professional learning, and assessment strategies. In curriculum-anchored assessment systems, the professional learning tied to assessments (e.g., designing, using, and scoring performance tasks) can become a form of curriculum-based professional learning, which evidence suggests is a particularly high-leverage strategy for improving classroom practice. States should elevate this aspect of the system, ensuring that there are systemic opportunities for educators to develop an understanding of how to design and use assessments to support their specific curriculum contexts. This builds on evidence suggesting that curriculum-based professional learning is particularly impactful in building instructional practice for teachers and meaningful learning for students.[12]

This might look like:

- connecting state-provided professional learning for task development, use, and student work analysis to curriculum (e.g., the instructional shifts governing many high-quality instructional materials, the instructional model within high-quality instructional materials);
- coordinating a vetted network of professional learning providers who offer specific curriculum-anchored professional learning connected to state-provided assessment resources;
- partnering directly with a specific technical assistance partner or partners to offer development and scoring workshops; and
- partnering with curriculum-related professional learning partners to sponsor and support professional learning communities to integrate assessment into curriculum-based professional learning.

Educative

Assessments build educator understanding of effective teaching and how students learn in the discipline.

Why is this important?

A major concern about current assessment systems is that they are used to justify ineffective, sometimes harmful, teaching and learning practices.[13] In assessment systems designed to support better teaching and learning, assessments—including the tasks, student data, and released materials and supports—must be designed such that they (1) build educator understanding of what high-quality teaching, learning, assessment, and feedback cycles look like, and (2) are meaningful learning experiences for students unto themselves.

By designing assessments to be educative to educators and students alike, state assessments can be transformed from something done *to* teachers and students—an act that is often grounded in distrust of teachers and gotcha moments for students—into a tool used in service of growth, continuous improvement, and cultivating teacher capacity and agency. Assessments designed to include elements that build teacher understanding will improve instructional experiences and contribute to curriculum equity. Moreover, state and local leaders can use educative assessments as the foundation for high-quality professional learning that bridges the assessment to curriculum implementation. For example, educators within districts often note that assessment scores alone often drive short-term moves that do little to support student learning and retention, like adding a unit on the scientific method to address perceived student deficiencies in science and engineering practice or repeatedly practicing reading decontextualized passages to practice skills like finding the main idea. However, when teachers are invited to examine tasks and student work (e.g., as part of assessment development or scoring workshops) that are illustrative (often annotated explicitly) of what teaching and performance in science or reading should look like, they find that their discussions shift toward deeper understanding of state standards, needed shifts in practice, and concrete next steps.

> State and local leaders can use educative assessments as the foundation for high-quality professional learning that bridges the assessment to curriculum implementation.

This is a particularly important element of assessment systems that are seen as valuable to teachers and students, making it essential for instructionally relevant assessment systems to sustain and scale. (See Table 6.)

Table 6. Characteristics of Assessment Systems That Are Educative

Assessment systems that are educative are less likely to ...	Assessment systems that are educative are more likely to ...
• Function as a black box for teachers and leaders (i.e., they do not get to see items or tasks or how students perform on the items and tasks that constitute the assessment). • Provide insights that lead to common misinterpretation of data or ineffective teaching practices (e.g., away from a coherent instructional model because the assessment does not distinguish between proficiency and the opportunity to learn). • Encourage implementation that reinforces ineffective teaching practices within the discipline. • Emphasize assessments that are not a learning experience for students. • Provide data that completely overlap with what teachers already know about student learning. • Confirm ineffective/incorrect ideas about student learning.	• Give teachers access to items, tasks, and student work that students complete as part of assessments. • Involve teachers in the evaluation and scoring of meaningful student work. • Expose and give teachers opportunity to practice high-quality teaching approaches in the discipline as well as give students opportunity for meaningful learning through completing the tasks. • Provide insights that are a value-add over what teachers and districts already know (e.g., could be insight into data that teachers are not already privy to, such as student experience data in tandem with performance information).

Source: Learning Policy Institute. (2024).

Critical Features of This Design Principle

1. **Provide transparent access to high-quality tasks and student work.** Perhaps the single most important element of educative assessments is that educators have transparent access to high-quality tasks and student work from the state assessment. It is nearly impossible to use assessments to improve practice based on numerical scores alone. When teachers and leaders can see what students were asked and how they responded, they can (1) make more principled decisions about how to interpret and use reported data, (2) understand what it looks like for students to demonstrate targeted learning in context, (3) connect student performance back to their own teaching practice, and (4) notice nuances in student thinking and performance to help pinpoint needed support for individual students and groups of learners.

2. **Center tasks that are illustrative of high-quality teaching.** Tasks used as part of state assessment systems should themselves reflect what high-quality disciplinary teaching looks like. For example, this may include:

   - providing guidance for appropriate customization of tasks (e.g., use of a more local example, use of more relevant texts or contexts),
   - providing opportunities for independent and collaborative thinking,
   - ensuring that assessment tasks follow a coherent storyline from the student's perspective,
   - offering students ways to build on prior knowledge,
   - offering students ways to engage in choice and decision-making within the task,
   - providing students with opportunities to iterate and revise thinking as new information emerges,
   - focusing on students using thinking and conceptual understanding over vocabulary use, and
   - including rubrics that are designed such that they help educators and students understand a progression of learning as well as highlighting useful feedback to support student growth.

   In curriculum-embedded performance tasks, this may look like providing comprehensive support for implementing the task. Tasks that do this may look quite similar to high-quality lessons and units and include supports for classroom discussion, links to build teachers' understanding of both the subject matter and discipline-specific pedagogical knowledge, and support for high-leverage disciplinary teaching practices appropriate to the nature of the task.

   In on-demand tasks that are released or otherwise made available for educators and leaders, this may look like detailed annotations that focus on associated teaching and learning practices (as opposed to only highlighting alignment) with supporting resources linked.

3. **Ensure that tasks provide meaningful learning experiences for students.** When done well, state assessment systems can design assessments such that students experience the assessment as a meaningful learning experience in its own right. To do so, state assessments might:

   - prioritize authentic, compelling phenomena and problems that involve true uncertainty and clearly matter to an authentic interest holder (rather than a generic "scientist" or "student");
   - leverage tasks that can surface what students know and can do in service of figuring out something relevant and new to them;
   - provide opportunities for students to reflect on their own learning and progress and reflect on how this task helped build and monitor that learning;
   - include opportunities for student choice and decision-making within tasks (e.g., selection among a set of comparable tasks or choices within the task around data or reading selections, arguments to be made);

- offer opportunities for students' own ideas to be an important element of engaging with the task (e.g., when making a recommendation for a design solution);
- provide opportunities to transfer understanding in ways that further build students' schema within a particular area; and
- provide educators with guidance, including professional learning and routines for implementation that support maximizing the learning experience for students.

4. **Provide support materials that connect assessments to high-quality teaching and learning.** States should prioritize making supportive resources publicly available, including task annotations and related resources, student work samples, and guidance for instructional and programmatic next steps. This may include links to external resources about the phenomenon or problem such that teachers could use these resources as part of classroom activities, connections to high-quality instructional materials that are publicly available, or recommendations to high-leverage practices that could be used at the school or district level, for example.

5. **Involve educators in assessment development, evaluation, and scoring efforts.** As discussed previously, one of the most powerful elements of instructionally relevant assessment systems is the connection to high-quality teaching and learning via professional learning. Developing systematic ways to involve educators in state assessment design, evaluation, and scoring efforts—and designing those efforts such that they provide educators with opportunities for collaborative sensemaking about tasks, student work, and connections to policies and practices that influence instruction—is essential for assessment systems to realize their potential to support teaching and learning. Note that this is predicated on the use of high-quality performance tasks as a central element of a given assessment. Many states try to employ a similar strategy in current assessment systems without compelling impact on instruction, due to the nature of the items on current assessments and the disconnect between those assessments and high-quality teaching and learning.

Developmental and Asset-Oriented

Assessments recognize what students *do* know and *can* do and surface progress relative to students' own performance and along appropriate learning progressions.

Why is this important?

Educators tell us that a major concern with current state assessment efforts is that while they produce a great deal of data describing and labeling whether students have met grade-level standards or not, they do not provide useful information about what students *do* know and *can* do in ways that can be used to make better decisions about instructional practice, curriculum supports, professional learning processes, and more. These data-rich but information-poor systems often encourage and reinforce ranking, labeling, and shaming students, teachers, and schools rather than pointing to productive next steps in service of growth and acceleration. Not only does this approach to assessment provide limited support for teaching and learning, but it also contributes to static, deficit-oriented narratives about students' disciplinary identities (e.g., "Student A is bad at science," "I'm not a science person") rather than more accurate and dynamic views that some students have simply not yet mastered some disciplinary learning goals and need additional support and opportunities to do so.[14]

If states instead focus on assessments that seek to identify more precisely what all students do know and can do, they can provide teachers, students, and families with assessment information that empowers them to take productive next steps to support growth. Research on how students learn makes it clear that students learn by building on prior understanding and experience, rather than simply accumulating isolated pieces of knowledge and skill. Assessments that are designed to surface facets of student understanding along within-year and multiyear learning progressions for disciplinary content and practices embedded in state standards can help students, families, educators, and leaders better identify assets of student thinking that can be leveraged as part of future learning opportunities. (See Table 7.)

Table 7. Characteristics of Assessment Systems Designed to Attend to Student Assets and Development Over Time

Assessment systems that are designed to attend to student assets and development over time are less likely to ...	Assessment systems that are designed to attend to student assets and development over time are more likely to ...
• Be used to label students and focus primarily on what they cannot yet do. • Overinflate what students know and can do or obscure learning edges. • Expect inappropriate levels of prior knowledge for students' developmental stages (e.g., expectations of transfer in science, cold reads in reading). • Use norm-referenced performance curves to map student progressions. • Yield data that show only whether students can perform at grade level, without information about how to meet students within their zone of proximal development. • Operate from a deficit orientation about what students are capable of doing (e.g., students are not reading at grade level, so we will limit the level of science thinking we ask of them instead of focusing on how to make the information within the text accessible to learners).	• Provide information about what students do know and can do, including relative to within-year and multiyear learning progressions. • Report growth and progress relative to evidence- and standards-based progressions. • Provide transparent information about student progress that can be used by parents, teachers, and students, grounded in the belief that students can accomplish high levels of work (e.g., provide guidance for scaffolding). • Provide appropriate accommodations and supports for students to show higher-order thinking.

Source: Learning Policy Institute. (2024).

What Does It Mean for Assessments to Be Asset-Oriented?

Imagine the following situation. On an assessment task, students are asked to make an evidence-based claim about what caused a change in a local ecosystem, using patterns in data that have been provided. A student writes "hunters and climate change." What can we make of this answer?

A deficit-oriented interpretation might focus on takeaways like "It's wrong," "The student doesn't get it," or "The answer isn't even related to the question." An asset-oriented interpretation of the same student response might recognize, "The student is making a claim that is relevant to the problem posed—they need some additional support in connecting this idea to the data provided," or "The student may be giving us ideas about their own experiences and interests—I wonder how we can lean into those experiences as a bridge to deeper understanding here."

While the student clearly needs more support and opportunity to develop the targeted understanding and skill, rubrics and feedback that focus on a deficit-oriented narrative are not productive. Instead, asset-oriented feedback focuses on noticing relevant facets of student performance that indicate what students know and can do at their learning edge, and that students and educators can build upon to support growth in learning toward the performance goals. Asset-based assessment design, reporting, and use operate from a belief that all students can be successful if given the right opportunities. They are honest and transparent about what students know and can do, while focused on what to do next rather than making judgments about students' current ability.

Critical Features of This Design Principle

1. **Assess learning along learning progressions.** Instructionally relevant assessments should assess understanding along learning progressions,[15] helping to identify facets of student thinking that educators can build upon. Because most states' standards are built upon intentional learning progressions that spiral across K–12, assessing along multiyear learning progressions can provide more accurate and asset-based accounts of what students know and can do, while still providing information that can be used to identify performance relative to grade-level expectations as required for state summative assessments under federal law.

2. **Use assessment tasks that provide all students with opportunities to feel successful by showcasing what they know and can do.** No student should leave an assessment feeling as though they could not demonstrate any understanding—this is both demoralizing for students and unhelpful to educators and leaders. Instructionally relevant state assessments should provide all students with opportunities to successfully make their thinking visible. This may be accomplished through a combination of on-demand items that range in complexity of sensemaking required as well as through compelling open-ended items and tasks that invite a range of thinking and response sophistication as students engage in sensemaking. Authentic, open-ended tasks in particular, when used in conjunction with asset-oriented student work evaluation and scoring processes, allow students to make more than simply a right or wrong answer visible. This is essential for students who may not have had access to high-quality academic instruction to still demonstrate assets in thinking that can be leveraged for next steps.

3. **Ensure that assessments reflect how learning happens in each discipline, not just mastery of end-of-instruction expectations.** If assessments are to operate from an assumption that students who are not yet proficient have not had sufficient opportunity to learn, rather than that they are just somehow fundamentally incapable, we must assess from an assumption of learning, not mastery. Students do not develop mastery of specific end-of-instruction standards by simply repeatedly "doing" the performance expected by the standard. For example, in science, students develop proficiency by using appropriate components of science disciplinary ideas, practices, and crosscutting concepts. In reading, students develop increasingly sophisticated text comprehension by iteratively connecting content knowledge and explicit reading skills. In all disciplines, students learn by being able to meaningfully engage with multiple representations of key content and practice, with appropriate scaffolds and opportunities for reflection and metacognition, over time and with increasing sophistication to support their use in meaning-making.

   For assessments to surface student thinking along a wide range of proficiency levels, it is important that assessments reflect how students progress toward mastery within each discipline. This may mean that assessments:

   - move away from complexity frameworks that position recall as the simplest and/or least sophisticated performance;
   - leverage features such as coherence from the student perspective, scaffolding/support within tasks, and attention to learning contexts to better understand the degree of transfer expected by assessments to vary the complexity of assessment tasks;
   - redefine proficiency levels (e.g., achievement-level descriptors and proficiency-level descriptors) to better attend to how learning happens in each discipline; and
   - make more appropriate and realistic claims about the degree to which students can transfer and generalize their knowledge within disciplines at different stages of development (e.g., 3rd-graders are unlikely to have sufficient schema or lived experience to be able to demonstrate their learning in contexts that are extremely different than ones in which they experienced learning).

4. **Provide rubrics, scoring processes, and reports that highlight use of asset-oriented narratives of students that focus on facets of student understanding.** Instructionally relevant state assessment efforts should make available information about student progress that encourages growth-oriented feedback and next steps. Rubric and performance-level descriptors should be precise about what students know and can do and avoid highlighting gaps in student achievement as the primary way of describing performance.

Reflective of and Responsive to Learners

Assessments follow principles of universal design and cultural responsiveness to ensure that each learner is supported in making their thinking visible.

Why is this important?

All students have and are part of rich cultural traditions that govern how they learn, understand, and make thinking visible.[16] Traditional approaches to state assessments that focus on standardization and "neutral" assessment contexts do not acknowledge that students' interaction with content is dependent

on lived experiences and how they developed their knowledge and skills—in other words, the sociocultural and linguistic structures surrounding the development of their disciplinary understanding. This means that many current state assessments are not measuring what they intend to, inadvertently reporting on students' language and cultural familiarity rather than their understanding and practice.[17] While states have generally taken more care to attend to some features of universal design, such as using multiple modalities to present information about phenomena and problems, current state assessments do not often extend these features to other ways in which students engage with and respond to assessments, again limiting the degree to which the assessment is yielding trustworthy information about some students' learning. (See Table 8.)

Table 8. Characteristics of Assessment Systems Designed to Be Reflective of and Responsive to Learners

Assessment systems that are designed to be reflective of and responsive to learners are less likely to ...	Assessment systems that are designed to be reflective of, and responsive to, learners are more likely to ...
• Assume the same "average" experience. • Assume that every student interacts with assessment tasks in the same way (e.g., an emphasis on standardization as the pathway to fairness). • Be designed in ways that exclude the experiences of students of color or students in poverty because they are not "majority" experiences. • Focus on decontextualized contexts, or contexts that are irrelevant to the students engaged in the assessment. • Include opportunities for students to connect with lived experience as a superficial component of the assessment. • Use proper standard written English as the only way students receive information and make their thinking visible. • Use rubrics and scoring guides that support only a single way of knowing as "correct."	• Center many different experiences and assume that a wide range of students will interact with tasks. • Center the lived experiences of students who have been historically underserved by educational experiences. • Provide opportunities for students' lived experiences and perspectives to be a central and valued component of completing assessment tasks. • Diversify who is considered a knower and doer in the given discipline. • Use diverse languages and dialects within assessment tasks and encourage the use of home language and multiple forms of making thinking visible by students. • Use rubrics that focus on meaning-making as the central element, allowing for multiple cultural ways of knowing to be valued (and include specific guidance and examples to support the inclusion of multiple ways of knowing).

Source: Learning Policy Institute. (2024).

Critical Features of This Design Principle

1. **Assessments should act as windows and mirrors for students. Student performance on assessments is the result of student-task-context relationships.** No assessment task will have the exact same relationship to every student because students bring different experiences and assets to the table when interpreting and responding to a task. Instead, states and developers should seek to design assessment instruments that are in dynamic relationship with students—that is, some students will be closer to a given task while others will be farther away, and the goal of assessment design is to vary this experience such that (1) all students feel seen and represented within an assessment, and (2) all students are supported in engagement with tasks, contexts, and/or ideas that might be less familiar to them. Designing assessment instruments that do this requires that tasks:

    - use a range of modalities (e.g., text, images, video, simulations) and types of text and information provided (e.g., graphs, charts, bulleted lists, different kinds of written sources);
    - center a range of perspectives and representations of who is a disciplinary knower, doer, and thinker (e.g., family or elder accounts, explanations that reflect different worldviews and experiences) as part of the information provided to students to shape tasks;
    - allow varied modalities and sensemaking routines to be valued in student responses. To the extent possible (e.g., open-ended tasks) and appropriate, assessments should strive for this to be true within a given task or item;
    - focus on assessments eliciting productive (not traumatizing) affective responses, rather than aiming for neutrality; and
    - position assessments as learning opportunities—students should feel that they learned about someone's lived experience (related to their own community or someone else's) by engaging in the task.

2. **Use assessment tasks that connect to students' lived and learning experiences.** Assessments should provide opportunities to connect to student experiences through (1) the phenomenon/problem-based scenario and contextual information provided, (2) the items and student responses, and (3) reporting and feedback processes. These connections must be as authentic as possible to avoid well-intentioned essentializing, stereotyping, and inappropriate assumptions about what is important to students and communities. This may look like:

    - leveraging known experiences from curriculum and high-quality instructional materials (connecting to the Curriculum Anchored design principle);
    - conducting student and community interest surveys and focus groups to determine relevant contexts and tasks;
    - providing choice to students among carefully designed tasks that measure the same construct but provide a range of contexts and ways to do so; and

- emphasizing tasks that leverage robust and common rubrics to allow students and educators flexibility around determining the specific elements of the task (e.g., students may be asked to design and justify a set of investigations to respond to a scientific question or area of inquiry but have flexibility in determining the exact nature of the experimental design, data to be collected, and implications for interpretation).

It should be noted that "students" and "communities" are not static, monolithic entities—both are dynamic and ever-evolving. This suggests that processes that seek to center student and community voice and experience should be approached as an ongoing dialogue rather than something undertaken once in time during an assessment development process.

3. **Position tasks to productively explore authentic and legitimate phenomena and problems that matter to specific communities.** While it may be challenging for states to generate tasks that are inherently interesting to each student engaged with an assessment, state assessments should strive to center tasks that are meaningful to a range of specific communities, such that the assessment as a whole is designed to be relevant to a true diversity of students, communities, and lived experiences. Tasks should position all communities represented as (1) more than stereotyped or victim-centered experiences, and (2) powerful doers and contributors to the discipline and the broader world.

4. **Centrally include tasks that authentically allow and encourage multiple perspectives and ideas as core elements of disciplinary reasoning and meaning-making.** Assessments should include tasks that (1) reflect a range of disciplinary ways of knowing (e.g., that there are multiple ways to "do science"); (2) position students' own ideas and perspectives—reflective of a range of individual and cultural identities—as an important element of disciplinary reasoning (e.g., when considering the trade-offs for possible design solutions); and (3) encourage and provide opportunities, connected to the disciplinary expectations, to critically consider what factors (historical, contemporary, social, cultural, environmental) have led to conditions and/or evidence being explored in a given task.

Useful for Informing Decisions That Impact Instruction

Assessments are designed to produce relevant information at appropriate times to support decision-making.

Why is this important?

Current assessments are often criticized for how irrelevant they are to instructional decisions. For example, data on current state assessments often do not become available to students, teachers, and leaders until the next school year is well underway, making it challenging to use current state assessments to inform even longer-term instructional planning and decision-making. Given the cadence of testing, this is particularly challenging in science—by the time teachers receive assessment scores, their students are often not even in the same building anymore, having moved on to middle or high school, which have markedly different structures and supports for learning. Moreover, when the information does become available, the nature of the information—numeric scores with little guidance for their interpretation—makes it difficult to connect to meaningful instructional shifts. Together, these features of current assessments surface perhaps the most urgent concern about the instructional relevance of state assessments: that they are simply not able to influence instruction.

Research and practical experience from states make it clear that across interest holder groups, people often make decisions that are most proximate to the data they have access to, in terms of both timing and content. States should intentionally consider this aspect of data use when designing instructionally relevant assessments, asking the question "How can we provide interest holders with the right information at the right time to drive positive instructional shifts in the discipline?" (See Table 9.)

> States should ask, "How can we provide interest holders with the right information at the right time to drive positive instructional shifts in the discipline?"

Table 9. Characteristics of Assessments That Are Useful for Informing Decisions That Impact Instruction

Assessments that are useful for informing decisions that impact instruction are less likely to ...	Assessments that are useful for informing decisions that impact instruction are more likely to ...
• Provide decontextualized scores. • Yield red-yellow-green gap analysis summaries without attention to appropriate next steps for the students or the educators involved. • Produce information that reifies ineffective teaching and learning practices (e.g., drilling information out of learning sequence, remedial strategies). • Use subdomain-based subscores as the primary way of providing "actionable, instructionally useful" data. • Yield data that are presented without supports for making meaningful connections to instruction (e.g., planning, next steps).	• Provide contextualized information, not limited to numerical scores, that is educative to students, teachers, and families alike. • Produce information aligned to specific instructional shifts that the assessment is intended to support. • Produce information that can be used by decision-makers, coaches, and teachers to make discerning decisions about high-impact actions, such as instructional materials adoption, professional learning time and content, and instructional design. • Provide states with support to interpret data resulting from an assessment relative to classroom, school, and district activities. • Provide insights that are directly connected to desired instructional shifts.

Source: Learning Policy Institute. (2024).

Critical Features of This Design Principle

1. **Make specific and explicit claims about instructional impact, and design accordingly.** States often make claims about increasing the quality and utility of feedback for students and teachers when designing assessment systems—but fail to specify *what* would increase the quality and utility, *who* should be using it, and *what decisions and actions* those specific actors should be making. States

and assessment developers should make clear, explicit, and discerning claims about what specific instructional impacts they want their assessment to produce. This should also include attention to unintended consequences of actual use and how the assessment is designed to protect against those unintended consequences that might produce a negative impact on instruction.

2. **Make available data that are closest to the desired instructional shifts.** States should carefully consider the instructional shifts they want to drive and identify how to make compelling data available through the assessment system to support that shift. This may look like:

   - making available different information on student reports, such as redesigning score reports to include information about opportunity to learn to contextualize proficiency/achievement data—including anonymized/aggregated information on student interest, experiences, and opportunity to learn to accompany score reports—or including annotated samples of assessment tasks and student work to accompany performance/achievement level descriptors and reports of student performance;

   - presenting data in new ways, such as exploring different reporting categories connected to instructional shifts or including information about student highlights and competencies mastered rather than scores and performance levels alone; and

   - focusing on a broader range of information that goes beyond scores and subscores, such as sample tasks, student work, analysis of what students' response patterns say about how they are thinking and needed additional supports, and descriptive callouts about student performance on specific aspects of an assessment (e.g., the performance assessments vs. scenario-based item clusters).

3. **Design assessment systems so that their content and timing align with intended and likely use.** While it may seem like more frequent information (e.g., via through-year assessments administered three or more times per year) will necessarily be more instructionally useful, this is not always the case. External assessments nearly always cause some disruption to teaching and learning, so assessments designed to be administered more frequently must be worth the disruption and potential costs to teaching and learning. Additionally, assessments administered more frequently have the potential to exacerbate the harms end-of-year testing already does in terms of reinforcing ineffective teaching and learning.[18] Indeed, when asked about current needs in assessments, local leaders frequently raise concerns about the quality and use of currently available interim assessments, which often derail well-designed and evidence-based instructional plans and models. It will be essential that states do not inadvertently position state assessments (e.g., through-year designs) to do the same.

   As states consider the timing and cadence of assessments they use and make available, it will be especially important to consider how measuring different aspects of disciplinary standards influences how educators should make sense of resulting information. For example, science and engineering practices and crosscutting concepts in science standards are expected to build over the course of entire grades and grade bands, while specific disciplinary ideas may be addressed in more modular ways. An assessment administered earlier in an instructional progression would have to be sensitive to these differences and help educators understand how to interpret the resulting student work.

Systems in Progress: Assessment Models and Examples

Operationalizing these design principles requires that states consider their specific intended impacts and needed shifts; current and historical culture around teaching, learning, and the state role in curriculum guidance; and the shifts needed within their own state education agency context. There is no single right way to implement these principles—states and assessment designers might use many different starting points and many different components of their assessment system design to begin redesigning for change, depending on the current state of their assessment and the desired end goal. Table 10 describes some possible design options that states could pursue.

Table 10. Examples Connecting Intended Action With Assessment System Design

If a state seeks to...	It might consider designs that...	Relevant examples
Increase uptake of high-quality instructional materials (HQIM)	• Explicitly leverage specific elements of targeted HQIM in assessment design • Provide flexibilities for approved performance tasks that are part of HQIM to be used as part of a state's assessment system	• Using anchor phenomena, contexts, or texts that are present in one or more HQIM as part of some tasks on a state assessment • Using the instructional model of one or more HQIM to guide the storyline of a state assessment performance task, used in conjunction with a sit-down assessment • Allow student performance on HQIM-embedded transfer tasks to be used in lieu of a state-provided performance assessment option (e.g., to meet a local performance assessment requirement, as part of a state's summative assessment score)
Improve HQIM implementation, including appropriate local adaptation	• Are clearly connected to HQIM (e.g., through scope and sequence, text, phenomena, and problem selection) but provide specific additional information critical to implementation • Include well-designed periodic assessments that are very coherent with specific instructional materials • Include the opportunity to learn and experience data to contextualize student performance	• Provide tasks that leverage related phenomena and problems that are intentionally designed to measure transfer and/or generalizability (i.e., by accounting for what students experienced in the classroom) • Use through-year assessment designs that are designed around specific high-leverage attributes of student thinking to help pinpoint specific instructional needs as they move through storyline units • Provide support for curriculum-embedded use of exit tickets and student interest and/or experience questions to help modify and support instruction appropriately • Provide curriculum-embedded tasks that leverage a common theory of learning with HQIM in science (e.g., sociocultural) but expand the materials to include extended projects that offer students ways to engage their families and communities, explore multiple worldviews, engage in disciplinary writing, explore local phenomena and contexts, and more

If a state seeks to...	It might consider designs that...	Relevant examples
Support curriculum-based assessment literacy and professional learning	• Use curriculum-embedded performance tasks that are designed to work with one or more HQIM • Prioritize educator engagement with student work	• Allow student performance on embedded transfer tasks in HQIM to be used in lieu of a state-provided performance assessment option (e.g., to meet a local performance assessment requirement, as part of a state's summative assessment score) and offer statewide student work analysis and scoring workshops • Use state performance assessments aligned to multiple HQIM to bridge state assessment and curriculum implementation in professional learning • Use teachers to audit AI-scored performance tasks
Incentivize specific instructional practices, like meaningful engagement with disciplinary practices, supporting student discourse within the discipline, or engaging students in complex problems and projects	• Include exemplary curriculum-embedded performance tasks (aligned to specific instructional shifts) • Release exemplar on-demand performance tasks routinely • Provide access to tools students have available on the test that can support teaching and learning as well	• Consider portfolio options, connected to instructional shifts, for a component of a state's assessment system • Create exemplar tasks, aligned to the instructional models of multiple HQIM, that are required as part of the state's assessment system and align with or replace curriculum-based transfer tasks or components of units
Support student agency and heterogeneous student outcomes and pathways	• Allow student choice in a component of the assessment system • Provide contextualized and diversified information about student assets that could support next steps toward a range of outcomes • Prioritize reports that highlight strengths and competency development	• Include student-led capstone projects and presentations as part of a multiple measures assessment system, used for consequential decisions • Provide opportunities for students to leverage validated authentic work done within school or outside of school as part of their demonstration of competency • Consider competency-based reports and transcripts that allow students to reflect a wider range of learning and success goals
Support the cultivation of teacher agency and reprofessionalization	• Provide teachers with supported choice about which tasks to use, and when to use them within the curriculum, for curriculum-embedded assessment opportunities • Involve teachers in assessment design and scoring processes • Focus on rich educative materials and tasks as a central component of state assessment efforts • Approach student assessment as a collaborative effort between teachers and the state	• Consider state assessment designs that are a combination of tasks developed and scored by teachers and tasks developed and scored by the state, with appropriate vetting and auditing as needed (e.g., IB assessment systems) • Leverage teacher scoring and AI scoring together for efficient scoring of tasks that do not sacrifice the potential for educator capacity building • Center performance assessment implementation and scoring on teacher collaboration within buildings and across sites • Emphasize teachers' learning edges—identified by educators themselves—in professional learning associated with the assessment (e.g., culturally relevant assessment practices, digital skills and literacy, effective student discourse)

If a state seeks to...	It might consider designs that...	Relevant examples
Improve the quality and utility of the statewide summative assessment	• Include rich, educative tasks and materials that significantly contribute to student scores and are released routinely • Emphasize teacher scoring of assessment tasks as a central element of the assessment design • Focus on developmental paths through learning, such that the data schools and teachers receive point to specific next steps • Provide information about student experience to directly contextualize performance measures	• Host statewide student work analysis workshops • Recruit rotating teacher panels for centralized scoring • Provide score reports that include examples of activities students can complete and suggested lessons/instructional resources to support growth • Include aggregated student experience survey data on student score reports

Source: Learning Policy Institute. (2024).

While most states in the United States are still exploring options related to instructional relevance, there are many positive examples emerging within (1) states; (2) systems that operate across many schools, districts, states, and countries, like Advanced Placement (AP) and International Baccalaureate (IB); (3) large school districts; and (4) internationally. These examples, operating at the intersection of multiple design principles, include systems that are:

- reimagining what—and how—to assess in large-scale assessment systems,
- centering teacher growth and learning via assessment systems, and
- changing how students see themselves and make their thinking visible in large-scale assessments.

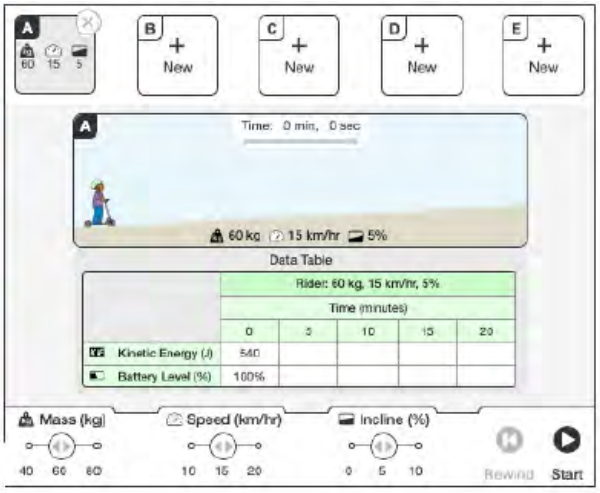
While still evolving, the assessments discussed here represent promising examples of how states and other large-scale systems can begin moving toward instructional relevance in intentional and meaningful ways.

Systems That Are Reimagining What—and How—to Assess in Large-Scale Assessment Systems

The following examples describe efforts from current large-scale systems to change the nature of the items and tasks used to measure student progress and proficiency.

Massachusetts Department of Elementary and Secondary Education (DESE) has developed an innovative science assessment system that connects on-demand, simulation-based sensemaking tasks (Figure 2) with rich classroom performance assessments (Figure 3) to provide teachers with better information about what students know and can do than via a selected response assessment alone.

Figure 2. Snapshot of a Sample Summative Assessment Task



This question has three parts.

Click here to learn how to use the simulation.

Part A

Identify **one** form of energy the battery's chemical potential energy was converted into when Maya rode her scooter to the tech museum. Explain how you know the energy conversion took place.

B I U 1500

Part B

Explain why the battery on Samuel's scooter had less energy than the battery on Maya's scooter after they traveled together to the tech museum.

B I U 1500

Part C

After the tech museum, Maya and Samuel plan to go to either the library or a store. The table shows the distance and incline of the two routes and of their original route to the tech museum.

Route	Beginning Battery Level (%)	Distance (km)	Incline of Route (%)
scooter station to tech museum	100	5	3
tech museum to library	100	5	5
tech museum to store	100	5	10

Identify whether Samuel would be **more likely** to reach the library or the store without using all the charge in the battery. Explain your answer using data from the table.

B I U 1500

Note: In this example, students are asked to use a simulation to understand why two different scooters might use energy in different ways. In the first part of the task (not pictured), students explore relationships among mass, speed, and force to make sense of different uses of the scooter battery. Later in the task, students are asked to consider how energy use is related to distance and incline and make predictions about how far the riders could go without using up all of their battery.

Source: Massachusetts Department of Elementary and Secondary Education. *STE Pilot Research Center: Innovative Assessment Program*. https://ma-innov-sci.mypearsonsupport.com/practice-tests/

Figure 3. Snapshot of a Sample Classroom Performance Task

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.

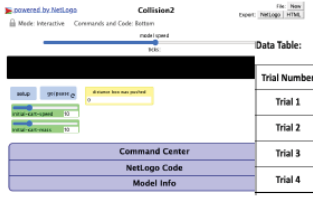
After reading short introductory material about the phenomenon and watching a short video, students are asked to complete three sets of activities:

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**.



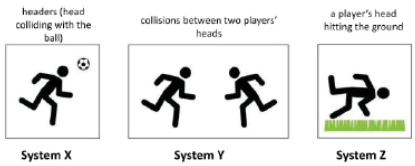
Data Table:

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**?

Note: In this classroom performance task, students individually and collaboratively investigate a relevant, real-world problem, trying to understand and address the prevalence of concussions in youth soccer. Through the task, students develop their own ideas, engage in social meaning-making, conduct investigations, make sense of data, develop and revise models, and construct arguments. The task gives students practice with simulations and analysis in classroom contexts that are similar in some respects to the summative assessment, but distinct enough to more fully be situated in classroom practice.

Source: Massachusetts Department of Elementary and Secondary Education. *Massachusetts Comprehensive Assessment System: Local-level classroom STE performance assessment tasks.* https://www.doe.mass.edu/mcas/tdd/ste-assess-pilot/classroom-tasks.html

Massachusetts's system design uses the idea of "bridging" to connect the performance tasks to the on-demand simulation-based tasks. The performance tasks themselves are designed to mirror features of the on-demand tasks that educators have identified as needing to practice with students (e.g., include simulations, follow a storyline) while also including features that connect the performance tasks directly to high-quality curriculum (e.g., guidance for how to use the task with multiple high-quality instructional materials, including modifications of transfer tasks directly from high-quality materials). While the classroom-embedded performance assessments are not currently scored as part of the test, they are designed and communicated by the state education agency (SEA) as being a critical element of preparation for the end-of-grade-band test.

Key features to note in Massachusetts's science assessment system include the following:

- All performance tasks are designed to reflect how students learn and demonstrate knowledge in science. They include collaborative and individual components, progressive and supported opportunities to make sense of phenomena and problems, and opportunities to make connections between lived experiences and the authentic situation presented in the task—and these are all central to how students make their understanding of science ideas and practices visible.
- The SEA began investing in classroom-based performance tasks in part because of the unintended impact the innovative summative test was having on instruction. Even with more careful attention paid to increase the authenticity of the on-demand test, SEA staff heard during piloting that the test was still being used to promote more limited and/or ineffective instructional practice in science. For example, teachers were spending time asking students to practice simulations rather than authentic engagement with meaningful phenomena and were limiting engagement with science practices like modeling and argument to mirror test-based proxies for these rich approaches to sensemaking (e.g., focusing on labeling diagrams instead of developing models to make sense of phenomena).
- The performance tasks provided the SEA with an opportunity to consider connection to curriculum more explicitly in both assessment design and the professional learning that surrounded the assessment system. The new performance tasks include an explicit design connection to high-quality instructional practices and materials. Some tasks are directly modified from high-quality instructional materials like OpenSciEd, with guidance for how to use the tasks with a range of other materials. Other tasks have been developed specifically for this assessment system but with careful attention paid to (1) how standards are unpacked and bundled in curriculum, and (2) mirroring the instructional models, practices, and routines that are present in high-quality instructional settings as the mechanisms by which students engage in the performance tasks (e.g., attention to collaborative investigation and sensemaking routines).

In **AP Computer Science Principles**, a substantial part of students' AP score for the course (used to determine whether students can receive college credit) is determined by the CREATE task, which asks students to develop a computer program—arguably the best indicator of whether students understand and can use the knowledge and skills the course seeks to develop. (See Figure 4.) The CREATE task is a curriculum-embedded opportunity to apply the skills being learned in the course. It seeks to be as authentic as possible: Students are encouraged to use the programming language of their choice, may collaborate with other students as well as artificial intelligence supports, and may develop a program that addresses goals of students' choice. Students are ultimately evaluated individually on how well they can describe and explain their program and how it was developed.

Figure 4. AP Computer Science Principles Classroom Performance Task

Create Performance Task

Programming is a collaborative and creative process that brings ideas to life through the development of software. In the Create performance task, you will design and implement a program that might solve a problem, enable innovation, explore personal interests, or express creativity. Your submission must include the elements listed in the Submission Requirements section below.

You are allowed to collaborate with your partner(s) on the development of the program only. **The written response and the video that you submit for this performance task must be completed individually, without any collaboration with your partner(s) or anyone else.** You can develop the code segments used in the written responses (parts 3b and 3c) with your partner(s) or on your own during the administration of the performance task.

Please note that once this performance task has been assigned as an assessment for submission to College Board, you are expected to complete the task without assistance from anyone except for your partner(s) and then only when developing the program code. You must follow the Guidelines for Completing the Create Performance Task section below.

General Requirements

You will be provided with a minimum of 12 hours of class time to complete and submit the following:

- **Final program code** (created independently or collaboratively)
- **A video that displays the running of your program and demonstrates functionality you developed** (created independently)
- **Written responses to all the prompts in the performance task** (created independently)

Scoring guidelines and instructions for submitting your performance task are available on the **AP Computer Science Principles Exam page** on AP Central.

Note: Students in nontraditional classroom environments should consult a school-based AP Coordinator for instructions.

Source: CollegeBoard. (2020). *AP Computer Science Principles: Course and exam description*. https://apcentral.collegeboard.org/media/pdf/ap-computer-science-principles-course-and-exam-description.pdf

Key features to note in the AP Computer Science Principles assessment include the following:

- By both requiring the completion of this performance task and connecting student performance on the task to summative scores, AP Computer Science Principles provides educators with an incentive to emphasize deep conceptual understanding and application within instruction, rather than emphasizing disconnected theory or rote and procedural skill without the accompanying understanding of how and when to use that skill.
- Emphasizing performance that is authentic to the discipline (i.e., programming an app) is more meaningful to students. In contrast, AP Computer Science A, which does not include performance assessments and emphasizes coding languages over application, experiences lower student retention in the course, less diverse students engaging with the course, and less positive student and teacher feedback on the course itself.
- While the performance task is a component of students' scores, it is experienced as a meaningful learning opportunity in and of itself. This reduces the assessment burden on teachers and students and limits the perceived trade-off educators must navigate around instructional time and assessment time.
- Despite being a well-known task by both students and educators, the validity of the assessment is not compromised—the assessment is valid *because* of the authenticity of the task. Prior knowledge of the task does not compromise student performance.

In **Louisiana**, the state is designing a curriculum-specific assessment in English language arts (ELA), currently being piloted under the Innovative Assessment Demonstration Authority. Teaching, learning, and assessment experts are collaborating to build a through-year assessment model in which students are asked to demonstrate key skills in reading and writing by engaging with texts that are familiar to students, either because they engaged with the texts directly in class (termed "hot reads") or because they are texts about a topic that students learned about through the course of instruction ("warm reads"). This assessment design leverages common knowledge that students are expected to develop through the curriculum, maximizing fairness as well as the interpretive value of assessment results. (See Figure 5.)

Figure 5. Sample Writing Task Leveraging Hot and Warm Reads

Directions and Sample Writing Tasks for Grade 5

You will now respond to a writing prompt based on what you learned during the unit and from the new text(s) you read in Section 1 of the assessment.

When responding to the prompt, be sure to do the following:

- **Organize and fully develop your ideas with thoughtful analysis.**
- **Include relevant information from the texts.**
- **Identify the texts you are referencing.**
- **Use correct grammar, punctuation, and spelling.**

Expository Task

The texts in the Making of a Scientist unit and in the excerpt from the unit-related text show how scientists can benefit from other people. Write a well-developed essay explaining how scientists benefit from other people in their lives and careers.

Support your explanation with ideas and information from the **unit-related excerpt** and from **one** of the unit texts listed below:

- "The Making of a Scientist" by Richard Feynman
- *The Templeton Twins Have An Idea: Book 1* by Ellis Weiner
- "Galileo Galilei: Biography, Inventions & Other Facts" by Nola Taylor Redd from SPACE.com
- *Giants of Science: Isaac Newton* Chapter 8 by Kathleen Krull

Opinion Task

The texts in the Making of a Scientist unit and in the excerpt from the unit-related text show how scientists can benefit from other people. Write a well-developed, multi-paragraph opinion essay about the most important benefit that scientists gain from family and/or other scientists.

Support your opinion with ideas and information from the **unit-related excerpt** and from **one** of the unit texts listed below:

- "The Making of a Scientist" by Richard Feynman
- *The Templeton Twins Have An Idea: Book 1* by Ellis Weiner
- "Galileo Galilei: Biography, Inventions & Other Facts" by Nola Taylor Redd from SPACE.com
- *Giants of Science: Isaac Newton* Chapter 8 by Kathleen Krull

Source: Louisiana Department of Education. (2024). *2023–2024 IAP assessment guide for grades 5–8 ELA guidebooks operational tests*. https://www.louisianabelieves.com/docs/default-source/assessment-guidance/leap-ela-guidebooks-iap-operational-assessment-guide---grades-5-6-7-8.pdf

Key features to note in Louisiana's ELA assessment:

- The model acknowledges that the ways students demonstrate key skills in the discipline (e.g., reading comprehension, synthesizing across texts) are closely tied to their knowledge about the topic at hand. When connected to the curriculum, the assessment can more fairly assess ELA skills because all students engaged in the same content.
- Because teachers know that the assessment is asking students to engage with material in the curriculum, there is an incentive to use high-quality instructional materials.
- In Louisiana, the connection between curriculum and assessment has allowed for a powerful and coherent professional learning model to emerge—all professional learning offered by the state is connected to the implementation of high-quality curriculum in the classroom.
- In focus groups, teachers have shared that the connection to curriculum makes assessment information (e.g., reports) much more useful to individual teachers as well as for conversations across classrooms, schools, and districts using common materials in the state.

CenterPoint is developing curriculum-connected performance assessments in science, focusing on bundles of standards that follow the scope and sequence of high-quality science instructional materials. In science, the assessment is designed to measure standards through multiple distinctly designed elements of the test: (1) curricular performance tasks that closely reflect the recently taught units, (2) longitudinal performance tasks that allow students to demonstrate how ideas and practices they are developing over time are progressing, and (3) standards-sampling items that are aligned to the scope and sequence of the curriculum units and are intended to more closely mirror other large-scale assessment instruments.

Key features to note in the CenterPoint science assessments include the following:

- When CenterPoint was developing its assessment system, it learned that pure alignment to standards was not sufficient for assessment tasks to feel consistent with high-quality curriculum. Other features, such as the nature of the phenomenon or problem addressed, the role of students within the task, and the nature of sensemaking, were essential for ensuring not only alignment to standards but also coherence between curriculum and assessment. The combination of all these elements was essential for teachers and students to consider the assessment a fair measure of what students know and can do.
- By anchoring to curriculum in different ways (e.g., close curricular tasks vs. longitudinal tasks), assessments can more explicitly identify the extent to which students can transfer and generalize their understanding of the disciplinary core ideas, science and engineering practices, and crosscutting concepts.

Systems That Are Centering Teacher Growth and Learning via Assessment Systems

In addition to states pursuing new assessment designs, several large-scale programs are also transforming how they leverage assessment instruments and results in service of more meaningful teacher growth and capacity building, focusing on the educative nature of their assessment systems. The examples below highlight systems that prioritize building teacher practice as part of the core purpose of their assessment endeavors.

Massachusetts DESE has transformed the professional learning it offers around its innovative science assessment system. By using the authentic performance tasks described above, DESE has begun reconceptualizing "assessment professional learning" to become about meaningful science teaching and learning processes. By centering the professional learning experience on task analysis, curriculum connection, and student work analysis, DESE has pivoted from a top-down information session *about* a test to a collaborative and generative professional learning model that centers using assessment to enhance instructional practice.

Smarter Balanced Assessment Consortium assessments in math and ELA include performance tasks that ask students to demonstrate bundles of disciplinary standards together in service of an authentic purpose as part of the summative tests. The Building Educators Assessment Literacy (BEAL) project examined how these tasks can be used to support teacher practice by designing teacher professional learning around understanding the design and use of Smarter Balanced performance tasks and directly evaluating and scoring student work. When teachers were able to see the performance tasks students

responded to along with student responses, and collaboratively score and discuss student work, the assessment transformed into a tool that directly built educator capacity, creating a substantial value-add for teachers. For example, 84% of teachers reported that they better understood their standards, and 92% of teachers had an understanding of how to use authentic tasks with their students in their classrooms.[19]

Hawai'i Department of Education (HIDOE) has launched the Performance Assessment Development Initiative (PADI) in math, ELA, science, and social studies. Through PADI, HIDOE is doing two important things: (1) signaling that rich, curriculum-embedded performance-based experiences that intentionally attend to students' cultural and linguistic assets are an important and valued element of meeting the state's adopted learning goals, and (2) providing an opportunity for rich professional learning for teachers and school leaders. In science, small cohorts of educators have routine opportunities throughout the year to collaboratively evaluate their curriculum, adapt performance assessments to meet key elements of culturally and linguistically relevant assessments, implement the tasks in their classrooms, and come back to discuss student work and next steps. As one teacher noted:

> Having an opportunity for ongoing support for curriculum-based assessment design and implementation has really helped me see how meaningful assessments can transform my actual classroom—not some abstract group of students at some time in the future, but my kids, right now. They have a much deeper understanding [of our content standards]. I'm excited to bring this learning to item development for the large-scale assessment!

The Washington Office of the Superintendent of Public Instruction has launched two related performance assessment initiatives centered on building teachers' science instructional practices: ClimeTime, which provides K–12 educators with access to authentic performance tasks focused on climate science, and the Science Assessment Grounded in Equity (SAGE) initiative, which provides a statewide set of resources, including vision, framework, assessment system, professional learning modules, and curriculum-embedded performance tasks designed to be more responsive to students and used as part of classroom learning in middle and high school. These initiatives have been coupled with (1) efforts to reduce the footprint of the state summative assessment in science; (2) statewide, ongoing professional learning support for assessment development, implementation, and use; and (3) considerations for how to use common features from the performance tasks as elements of the statewide summative assessment.

Systems That Are Changing How Students See Themselves and Make Their Thinking Visible in Large-Scale Assessments

Several large-scale assessment systems are exploring and implementing ways to make their instruments more reflective of the diversity of students they interact with and more responsive to the wealth of cultural and linguistic assets that underlie student learning and performance. In doing so, programs are seeking to ensure two things: (1) that assessments are truly surfacing trustworthy information about what students know and can do, and (2) that assessments are not experienced as dehumanizing and demeaning. This has been in response to several calls from the field to ensure that culturally relevant assessment development and implementation practices are prioritized in large-scale efforts.[20]

The National Assessment of Educational Progress (NAEP) has begun reimagining how to address cultural and linguistic representation and repertoires in large-scale assessment systems. In the **NAEP Reading Framework**[21] for assessments beginning in 2026, there is a considerable emphasis on attending to the interaction between the knowledge students have developed (through social, cultural, and linguistic interactions inside and outside of school) and how they read and make sense of texts. For example, the new NAEP reading assessments include:

- expanded range of text types and modalities;
- navigation support to help students make connections across texts, disciplinary subscales, and informational introductions to the text (often using video and audio elements) to help provide students with necessary background knowledge;
- attention to larger language structures and more varied linguistic repertoires and types of discourse, representing a wider range of students, communities, and interactions; and
- expanded disaggregation to better understand how students with varied lived experiences are able to demonstrate reading comprehension.

Similarly, the **NAEP Science Framework**[22] for assessments beginning in 2028 includes explicit attention to the range of lived experience, cultural and linguistic backgrounds, and engagement and motivation factors for learners engaged in the assessment. New NAEP science assessments will include:

- authentic phenomena and problems grounding all assessment tasks, with attention given to contexts that focus on legitimate interests of specific communities;
- attention to varying which communities and lived experiences are centered;
- sufficient background information, provided via multiple modalities, to support students in engaging with tasks that are further from their own lived experience;
- diversified representation of who is considered a knower, doer, and thinker in science; this includes specific attention to people and communities of color, with an emphasis on asset-based narratives;
- open-ended tasks that center multiple perspectives in service of responsible decision-making;
- both context and prompt language, as well as acceptable student responses, that will attend to a wider range of language and linguistic repertoires; and
- when appropriate, tasks that allow for critical examination, via the science practices and appropriate disciplinary ideas, of claims, data, explanations, and accounts for phenomena and problems.

The **Center for Measurement Justice**, in collaboration with **Curriculum Associates**, is developing models of culturally relevant and antiracist assessment items. These items consider different approaches to culturally relevant/antiracist assessment design, such as:

- centering the experiences of Black, Brown, and Indigenous people and communities within texts and contexts provided within items;

- providing students with choices among contexts to allow them to engage with the topic or experience that is most relevant and compelling to them; and
- positioning items and tasks as learning experiences for students and teachers about antiracism and/or the lived experiences of others from marginalized communities.

In Hawai'i, the **Kaiapuni Assessment of Education Outcomes** (KA'EO) is a Hawaiian language assessment for Hawaiian language arts, math, and science used for local and federal accountability purposes. The assessment, designed to center Hawaiian language and culture in assessment in ways that are coherent with instruction in Hawaiian immersion schools, is an interesting example of active cultural and linguistic reclamation as the context for more valid and instructionally useful assessment measures. While many assessment designs that center multiple linguistic assets account for home language (e.g., students who speak Spanish at home but are learning and testing in English), KA'EO specifically attends to culture and language as they are developed within the school and classroom curriculum and culture. All assessment development activities are done in purposeful collaboration with teachers from immersion schools across all of the islands, and KA'EO is designed with attention to specific features of local communities, variations in language, and approaches to how educational philosophy in classrooms emerges from and interacts with attention to Hawaiian culture—ultimately leading to an assessment that teachers, students, and families trust and can be used to advance the goals of Kaiapuni schools. KA'EO is aligned to state standards through the lens of Hawaiian culture and language.

System Conditions for Instructionally Relevant Assessment Systems

For instructionally relevant assessments to reach their potential, they must be designed and implemented within systems that share a coherent goal of improving instruction. Some key system conditions that are important for instructionally relevant assessment systems to take hold are:

- **A Central Role for Professional Learning.** Even the best-designed assessment instruments can have only limited instructional impact if teachers, leaders, students, and families are not supported in making good use of better tasks and reports. Any assessment system must be designed with instructional impact in mind to center meaningful professional learning. This means assessment instruments must be designed to answer the question, "How can this assessment be used as a meaningful part of sustained, curriculum-based professional learning that is directly connected to teachers' and leaders' practice?" and state assessment systems should emphasize—in resource allocation, timing, and design—that thoughtful approaches to professional learning are a non-negotiable element of instructionally impactful assessment systems.

- **Opportunity-Focused Local Decision-Making.** Assessments should be a tool for meaningful learning—and yet, in too many classrooms, schools, and buildings, assessments are used to deny students access to AP, IB, and other advanced courses, real-world projects and service, and even access to core subjects because of a focus on remediation. The decision-making processes that leverage assessments must be designed to create more meaningful learning and actively disrupt barriers to student opportunity.

- **State Education Agencies (SEAs) as Partners, Supporters, and Co-Conspirators.** In many state contexts, schools and districts view the SEA as a compliance officer—seen as a top-down decision-maker, creating the guardrails and requirements that local leaders and practitioners are beholden to. Often, teachers and local leaders distrust the state because of prior experience of state policies being done "to" them, rather than with the input of practitioners. This kind of relationship between state education agencies, local education agencies, and local communities is unproductive and often does not reflect the intent of SEA staff. Indeed, most state leaders have expressed that they deeply share the commitment to providing more meaningful learning experiences for students and are simply at a loss for how to do so more effectively, especially while navigating the challenging and competing priorities states have to contend with.

  Truly advancing instructionally relevant state assessment systems requires reimagining the role of the SEA into one of partnership with local communities. This has just as much to do with habits and mindsets as it does with actual programmatic work and decisions—how SEAs do the work matters just as much as what work they do. For example, the same performance assessment initiative might feel exciting and supportive of teachers' and leaders' goals in one context, and like another frustrating hoop to jump through and another assessment requirement being done to schools in another context—the degree to which the performance assessment initiative is seen as a responsive, field-led initiative is often the distinguishing factor.

Reimagined SEA leadership might look like inclusive and empathy-oriented design processes for new state policies and initiatives, strategic enabling of local innovation, active removal of unnecessary compliance-oriented barriers, and revisiting with practitioners what kinds of guidance are most impactful.[23] Most importantly, this is cultural change—both within SEAs and in the dynamics of local–state interactions. This kind of work is dynamic, is ongoing, and requires sustained commitment.

- **A Supportive Policy Context.** States need to be confident that efforts to develop instructionally relevant assessment systems that they plan to use as their federally mandated statewide assessments will have the support of the U.S. Department of Education (USED) as they move forward. This does not mean blind support for any system that is proposed, but it does mean that USED should make principled decisions about (1) how and when to expand the interpretation of quality criteria for state assessments; (2) what to prioritize in terms of funding, waivers, and other pathways to innovation; and (3) what current barriers to more instructionally impactful assessments could be reduced or eliminated without compromising an unwavering commitment to equity for all students.

- **Supportive Technology.** Advances in technology—including, but not limited to, generative AI and other automated scoring processes, knowledge and learning management systems, and data transfer technologies—can make assessment processes that are otherwise cumbersome or resource-intensive more practical, allowing systems to explore innovative practices more equitably. Importantly, states and assessment system designers should consider how we can leverage technology *in service of* instructionally impactful systems. This might mean (1) considering authentic and immersive simulations that can bring authenticity to students' assessment experiences, (2) considering how AI can be used in tandem with teacher-centered student work analysis to balance scoring large samples of student work with intentionally educative assessment designs, and (3) considering innovative technology solutions used to capture and transfer authentic student work.

Conclusion

By centering features of assessments that support better student learning experiences, teacher practice, and systematic supports and decision-making, we can create assessment systems that have a net positive impact on instruction. The design principles detailed here reflect ambitious but accomplishable goals for our assessment systems—and large-scale systems, including states as well as national and international programs, are already on the path to making this work a reality. As systems move forward, keeping "positive instructional impact" as the North Star and centering decisions on specific instructional shifts from the current state of teaching and learning that assessments should support can help system designers make the best decisions within their local contexts.

Endnotes

1. Stecher, B. (2014). Looking back: Performance assessment in an era of standards-based accountability. In L. Darling-Hammond & F. Adamson (Eds.), *Beyond the bubble test: How performance assessments support 21st century learning* (pp. 17–52). Jossey-Bass; Stecher, B., Barron, S., Kaganoff, T., & Goodwin, J. (1998). *The effects of standards-based assessment on classroom practices: Results of the 1996–97 RAND survey of Kentucky teachers of mathematics and writing*. National Center for Research on Evaluation, Standards, and Student Testing; Faxon-Mills, S., Hamilton, L. S., Rudnick, M., & Stecher, B. M. (2013). *New assessments, better instruction? Designing assessment systems to promote instructional improvement*. RAND Corporation. http://www.jstor.org/stable/10.7249/j.ctt5hhtkh

2. National Academies of Sciences, Engineering, and Medicine. (2018). *How people learn II: Learners, contexts, and cultures*. The National Academies Press. https://doi.org/10.17226/24783; National Research Council. (2001). *Knowing what students know: The science and design of educational assessment*. The National Academies Press. https://doi.org/10.17226/10019

3. Koedel, C., & Polikoff, M. (2017). Big bang for just a few bucks: The impact of math textbooks in California. *Evidence Speaks Reports*, *2*(5); Kane, T. J., Owens, A. M., Marinell, W. H., Thal D. R. C., & Staiger, O. (2016). *Teaching higher: Educators' perspectives on Common Core implementation*. Center for Education Policy Research, Harvard University; Kaufman, J. H., Doan, S., & Fernandez, M. (2021). *The rise of standards-aligned instructional materials for U.S. K–12 mathematics and English language arts instruction: Findings from the 2021 American Instructional Resources survey*. RAND Corporation; Council of Chief State School Officers. (2023). *High quality instructional materials and professional development network case study: Impact of the CCSSO IMPD Network*.

4. Council of Chief State School Officers. (2022). *High quality instructional materials and professional development network case study: Signaling and incentivizing quality*; Council of Chief State School Officers. (2023). *High quality instructional materials and professional development network case study: Impact of the CCSSO IMPD Network*.

5. Achieve. (2019). Task Annotation Project in Science: Sensemaking.

6. National Academies of Sciences, Engineering, and Medicine. (2018). *How people learn II: Learners, contexts, and cultures*. The National Academies Press. https://doi.org/10.17226/24783; National Research Council. (2001). *Knowing what students know: The science and design of educational assessment*. The National Academies Press. https://doi.org/10.17226/10019

7. It should be noted that while current state assessments may not intentionally or explicitly connect with curriculum, no assessment is actually agnostic or neutral to curriculum. All assessments operate from some assumed priorities about how students do or should experience disciplinary teaching and learning. The problem with current assessments is that they are often signaling and incentivizing (likely inadvertently) curricular approaches that are not yet coherent with the research-based vision for teaching and learning states have adopted through their state standards. By promoting curriculum-anchored assessment systems, we are simply suggesting that states acknowledge this relationship between state assessment and teaching and learning decisions, and more intentionally make decisions about the state assessment, knowing the influence it has.

8. Badrinarayan, A., & Steiner, D. (2023). *Positioning state assessment systems in service to teaching and learning: The role of high-quality curriculum in state assessment design*. Education First.

9. A growing body of research suggests that access to high-quality instructional materials, coupled with curriculum-based professional learning, is one of the most powerful factors in shifting teaching and learning practices. While decisions about curriculum and instructional materials adoption and implementation are generally left to districts and schools, an increasing number of states are committed to exploring ways to leverage state policies to incentivize the adoption and use of high-quality instructional materials. In this context, it is imperative that states carefully consider the relationship between instructional materials and state assessment systems.

10. Ruiz-Primo, M., Li, M., Wills, K., Giamellaro, M., Lan, M., … Sands, D. (2012). Developing and evaluating instructionally sensitive assessments in science. *Journal of Research in Science Teaching*, *49*, 691–712. https://doi.org/10.1002/tea.21030; D'Agostino, J., Welsh, M., & Corson, N. (2007). Instructional sensitivity of a state's standards-based assessment. *Educational Assessment*, *12*(1), 1–22. https://doi.org/10.1080/10627190709336945

11. This idea is based on thinking shared by Scott Marion in personal communication and reviews of these design principles.

12. Short, J. B., & Hirsch, S. (2022). *Transforming teaching through curriculum-based professional learning*. Corwin.

13. National Education Association. (2021). *Beyond the bubble: Americans want change on high stakes assessments*; National Education Association. (2021). *Principles for the future of assessment*; American Federation of Teachers Assessment Task Force. (2024). *Real solutions for improving assessment*; Armstrong, T. (2013). *15 reasons why standardized tests are problematic*. ASCD.

14. Papay, J. P., Murnane, R. J., & Willett, J. B. (2011). *How performance information affects human-capital investment decisions: The impact of test-score labels on educational outcomes* (NBER Working Paper w17120). National Bureau of Economic Research; Boaler, J. (2011). Changing students' lives through the de-tracking of urban mathematics classrooms. *Journal of Urban Mathematics Education*, *4*(1), 7–14; Carey, R. L. (2014). A cultural analysis of the achievement gap discourse: Challenging the language and labels used in the work of school reform. *Urban Education*, *49*(4), 440–468.

15. Learning progressions can mean different things in the learning sciences, standards, and practice. While there is always more to learn about how student thinking progresses, current state standards in math, ELA, and science are based upon an assumption that students build on previous years' learning in each subsequent grade level. High-quality curriculum decisions and design follow these progressions—in addition to being better practice for learning, following these kinds of explicit learning progressions also helps assessments better attend to what students have had the opportunity to learn. Importantly, learning progressions here refer to established research- and practice-based expectations for progressive student understanding (e.g., as represented in state standards across grade levels), not norm-referenced ranges of performance.

16. Nasir, N., Lee, C., Pea, R., & McKinney de Royston, M. (2020). *Handbook of the cultural foundations of learning*. https://doi.org/10.4324/9780203774977

17. Note: In a related project, the Learning Policy Institute is developing a framework to guide culturally responsive assessment practices for assessments at scale, built upon work happening currently in states. Please see this framework for additional detail.

18. The Center for Assessment has written extensively on the necessary considerations for through-year assessments, and a full collection of its writing on the topic can be found at nciea.org. Key considerations for through-year assessments are discussed in Dadey, N., Evans, C., & Lorie, W. (2023). *Through-year assessments: Ten key considerations*. National Center for the Improvement of Educational Assessment.

19. Arnold, J. (2016). *Making the most of performance tasks in summative assessment: Building educator assessment literacy in Oregon*. WestEd.

20. See, for example: Munyan-Penney, N., & Mehrotra, S. (2023). *Future of assessments: Centering equity and the lived experiences of students, families, and educators*. Education Trust. https://edtrust.org/resource/future-of-assessments-centering-equity-and-the-lived-experiences-of-students-families-and-educators/; 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

21. WestEd & Council of Chief State School Officers. (2021*). Reading framework for the 2026 National Assessment of Educational Progress*. National Assessment Governing Board & U.S. Department of Education.

22. WestEd. (2023). *Science framework for the 2028 national assessment of educational progress*. National Assessment Governing Board & U.S. Department of Education.

23. Kentucky's current assessment and accountability project has leveraged radically inclusive design processes, local laboratories of learning, and statewide opportunities for curriculum-based professional learning, to both position the state as a partner in local innovation and to provide statewide guidance and opportunity that all districts can be a part of.

About the Author

Aneesha Badrinarayan is the Director of State Performance Assessment Initiatives at the Learning Policy Institute (LPI), where she supports states in developing meaningful assessment systems that are 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 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. Prior to LPI, Badrinarayan served as the Director of Special Initiatives at Achieve, as a museum educator, and as 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 a BA in Biology from Cornell University.

LEARNING
POLICY
INSTITUTE

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

@LPI_Learning | 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.