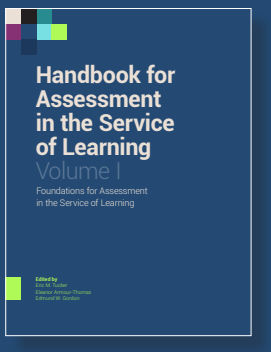
Reimagining State Assessments in Service of Teaching and Learning: Design Principles for Instructionally Relevant Assessments

Aneesha Badrinarayan

UMassAmherst
University Libraries



Series Editors:
Edmund W. Gordon, Stephen G. Sireci, Eleanor Armour-Thomas, Eva L. Baker, Howard T. Everson, and Eric M. Tucker

ISBN: 978-1-945764-33-2

Suggested Citation:
Badrinarayan, A. (2025). Reimagining state assessments in service of teaching and learning: Design principles for instructionally relevant assessments. In E. M. Tucker, E. Armour-Thomas, & E. W. Gordon (Eds.), *Handbook for assessment in the service of learning, Volume I: Foundations for assessment in the service of learning.* University of Massachusetts Amherst Libraries.

Reimagining State Assessments in Service of Teaching and Learning: Design Principles for Instructionally Relevant Assessments

Aneesha Badrinarayan

Abstract

This chapter presents a comprehensive framework for redesigning state assessment systems to better support instruction and student learning. Historically, large-scale assessments have prioritized easily quantifiable outcomes, often emphasizing surface-level skills over deeper, more meaningful learning. This approach has inadvertently shaped instructional practices, leading educators to focus on test preparation rather than fostering critical thinking and complex problem-solving skills. In response, this work proposes six key design principles to realign assessments with instructional goals and educational equity.

The proposed principles include: (1) Authenticity—creating assessments that reflect real-world disciplinary practices and engage students in meaningful tasks; (2) Curriculum-Anchoring—designing assessments that align with high-quality curricula to incentivize and support effective teaching practices; (3) Educative Design—using assessments to build educators' understanding of instructional quality and effective learning strategies; (4) Developmental and Asset-Oriented Approaches—recognizing students' progress along learning trajectories and focusing on their strengths; (5) Cultural Responsiveness and Universal Design—ensuring assessments are inclusive and reflective of diverse student backgrounds; and (6) Instructionally Useful Data—providing timely and actionable feedback to inform instructional decisions.

By implementing these principles, state assessments can become powerful tools for driving positive instructional change, supporting teacher development, and fostering equitable learning outcomes. This chapter calls for a paradigm shift in assessment design—one that balances rigor with relevance and centers the holistic development of every learner.

The decisions states and districts 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, large-scale assessments selected and administered by states and districts 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 performances 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 is little question that this is inconsistent with how assessments are actually used in practice (Badrinarayan 2024).

For example, consider the following reflections from high school teachers and university professors as they discuss how ELA assessments are influencing instruction, drawn from a focus group conversation as part of national focus groups conducted to better understand the relationship between teaching, learning, and assessment (State Performance Assessment Learning Community, 2022).

Professor: We're finding that students are admitted to our [college] programs based on high school grades and test scores that show some degree of proficiency in reading and writing…but students come to us completely unable to make sense of complex texts or write about them in deep, generative, thoughtful ways. One student came to my office hours last week in tears because she was overwhelmed by the reading load in my syllabus, which is just a few chapters of a couple of books each week. I was in utter shock when she told me she had not read a full book in the last two years of high school.

High School Teacher: That's not surprising to me. Our interim assessment cadence is really fast, and I need to keep kids moving through and give them lots of practice with the skills that are being measured, like finding the main idea. We don't

have time to read long books, so we do a lot of excerpts, short stories, that kind of thing, about a lot of different topics. But students get practice with a lot of different texts, so I think that's probably helping them get better and better at the skills I'm trying to teach them.

Professor: It seems like they are getting good at school and taking tests—this isn't what [ELA] is supposed to be about. Being a good reader and writer is about being a good thinker, a good maker of meaning. A skill like "finding the main idea" isn't something you just practice or develop in a vacuum—by the time students leave high school, they should be able to deal with nuance and complexity of negotiating meaning across multiple, robust texts, grounded in content and topical learning.

Teacher: Well, I have no idea what topics or texts students will have to read on the interim or state assessments, so I can't go really deep on any one topic or text. Those tests really focus on students being able to read any text about any topic, so it's less about bringing knowledge to the text, and more about just being able to decode and make meaning of the words on the page—the other stuff is for science and social studies to teach. The texts on tests are pretty simple and my students struggle with them as is, so I think the goals you're describing are a little too ambitious for most learners. Maybe for our advanced students! But so few of our students are close to "mastery" in terms of proficiency levels, I can't even imagine what the kind of performance you're describing would be—super mastery? What's the achievement level?

Professor: But that's because they're not actually engaged in meaningful comprehension activities if they just jump from text to text and skill to skill—that's not how you learn how to read and make meaning.

We can unpack this example—and many more like this—to pinpoint at least three key ways that state assessments influence the behavior of educators in ways that impact instruction:

1. **They signal both <u>what</u> and <u>how</u> to teach.** In the example above, the K–12 teacher notes specific cues around instruction that they are taking from systemic assessment experiences, such as intentionally emphasizing interaction with multiple, diverse, short texts over deeper sustained engagement with longer pieces, and practicing skills like finding the main idea over knowledge-building and related meaning-making of texts. This is consistent with decades-old adage

that "what gets tested gets taught," but highlights that this doesn't only happen at the level of content domains or major disciplinary subdomains—what is on the test influences what materials educators select to use with students, how they structure and scope student interaction with those materials, and what they prioritize in terms of practice- and knowledge-building.

2. **Tests signal the bar—and expected range—for student performance.** Large-scale assessments are usually developed with the expectation that they are measuring the "floor" of expectations for student performance. Like the standards they are intended to assess, they represent what all students should know and be able to do, without placing any limitations on teaching and learning that goes *beyond* the expectations of standards in developmentally appropriate ways. However, in practice, standards and assessments are often interpreted as truly reflective of the full range of expected instruction and performance. In the example above, the K–12 educator is clearly using the achievement or proficiency level descriptors associated with the state assessment as the markers of the expected range of student performance. Even when they are receiving direct feedback from a postsecondary interest holder about the need for something beyond what the test is measuring, the teacher assumes that this must be out of bounds of even what high-level students who have achieved grade-level mastery should accomplish.
3. **They reinforce orientations toward students and their potential.** The college professor opens this dialogue by stating a mismatch between what they believe students are *capable* of—reading and writing in deeply thoughtful and generative ways—and what students have been *given opportunity to develop* through school. However, the K–12 teacher explains this mismatch away by stating beliefs about what students are actually capable of doing, as evidenced by the test. "My students can't do that" is perhaps one of the most common reactions we hear when showing teachers examples of high-quality tasks designed to surface what students know and can do related to disciplinary standards and durable skills. When asked what changes need to be made for students to be successful, teachers often suggest changing the task, rather than the instruction leading up to it, to better reflect what students can possibly be successful with. These examples indicate that external, large-scale assessments reinforce beliefs that teachers have about what students have the potential to accomplish, creating a feedback loop that serves to shape instruction.

It would be disingenuous to suggest that tests have this kind of influence in a vacuum—state assessments likely have this influence *because* of how they have been positioned in the broader educational system. For example, student performance on state assessments is the overwhelmingly dominant factor in determining school accountability, which has historically had implications not only for school funding, but also teacher evaluations, school restructuring, and labeling of schools as "good" or "bad" by local communities and external parties. This positions state assessments as a central focus for school and district activities, and more decisions that impact instruction (e.g., instructional materials selection, scope and sequence determination, pedagogical emphases, and decisions about which students get access to different curricular experiences) are made in ways that align with signals from state tests. Similarly, because state assessments have been used in ways that are particularly consequential for adults, the specific technical tradeoffs that are made with regard to current large-scale instruments (e.g., emphasizing reliability and efficient scoring, prioritizing standards coverage over depth) are elevated as more robust and at a higher bar for quality than other instruments that make different, but equally reasonable, tradeoffs. These assessments and their technical qualities, then, are held up as a gold standard for other assessment design efforts, reinforcing their position as highly influential.

Together, these kinds of factors have created a culture in educational systems that values external, seemingly 'objective' assessment instruments and the data they yield. Even in subject areas where state assessments are much less consequential, such as science, we see similar behavioral patterns in response to state assessments from many different interest holders. Leveraging our understanding of how a range of users use state assessment information, assessments can be designed such that state assessments have a net-positive impact on instruction. We can do so by designing assessments (See Table 1) that:

- Are grounded in defined instructional shifts, from current practice, that assessments should be designed to incentivize and drive.
- 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) mirror 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 simply 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.

Table 1

Instructionally Relevant Decisions 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
Fairness with harmful bias managed	• 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

Table 1 (continued)

Interest holder	Examples of instructionally relevant decisions and actions
School and district leadership	• Scheduling decisions that enrich curricular opportunities for all learners (e.g., inclusive learning for emerging multilingual learners rather than pulling 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, assessment, 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: Badrinarayan (2025).

Assessments that do so acknowledge that, whether they intend to or not, large-scale assessments do impact instruction—facing this reality head-on means that these same assessments can be designed to contribute to more effective teaching and learning in both direct and indirect ways, rather than operating under the assumption that the ways large-scale assessments influence instruction are just "side effects" of tests designed to address other priorities.

The following design principles for large-scale assessments seek to position assessments that operate across many different schools and districts in ways that are in service of learning as described by the overarching ***Principles for Assessment in the Service of Learning*** (this volume). In particular, these design principles for large-scale assessments emphasize elements of the following Principles as those deemed (1) most appropriate for large-scale assessments, and (2) those most critically impacted by large-scale assessment design and use in both positive and unproductive ways:

> **Principle 1:** Assessment ***transparency*** provides clear information about assessment content and use to assist learners, teachers, administrators, and parents.
>
> **Principle 2:** Assessment ***focus*** is explicit and includes purposes, outcomes, progress indicators, and processes that can be transferred to other settings, situations, and conditions.
>
> **Principle 5:** ***Feedback***, adaptation, and other relevant instruction should be linked to assessment experiences.
>
> **Principles 7:** ***Assessment quality and validity*** should be available and reflect evidence related to assessment purpose to permit appropriate inferences and findings about quality, utility, and credibility.

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 (Badrinarayan, 2024; See Table 2.) 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 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; and
- walk the line between aspirational and doable—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.

This chapter provides an overview of the design principles documented by Badrinarayan (2024), including rationale for the design principle and critical features states and other large-scale assessment system designers may consider including in their assessment system to enact the principle within their systems.

This chapter uses the term "state assessment systems" to refer to the collection of assessment instruments, supporting documents, and assessment-related services 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-instruction assessment, or it may occur in a model that leverages multiple assessment opportunities throughout the course of an academic year or course of study (e.g., through-year assessments). In other states, this might encompass a combination of an on-demand statewide test with locally selected and/or used instruments (e.g., local performance assessments). The principles are intentionally focused on features of required 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 they deserve to be (Darling-Hammond & Adamson 2014, Darling-Hammond 2017).

Importantly, while these design principles have been developed to be ambitious, but achievable for states, they are strong indicators for potential 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).

Table 2

Design Principles for Instructionally Relevant Assessment Systems

Principle. Instructionally relevant assessment systems are intentionally designed to be…	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 representing 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. These tasks should also engage students in sensemaking in ways 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 designed such that high-quality curriculum better prepares students for success on the assessment, it incentivizes the adoption and use of high-quality curriculum materials, and it supports implementation of high-quality materials.	Assessments should signal, incentivize, and support the use of high-quality curricula 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 particularly useful to instruction; to encourage the use of instructional materials and models focusing on deeper learning; and provide students with meaningful learning experiences. In some cases, assessments may be designed to reflect high-quality instructional materials (HQIM); in other cases, they 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).

Table 2 (continued)

Principle. Instructionally relevant assessment systems are intentionally designed to be…	Summary Statement
Educative. Assessments build educator and student understanding of and experience with high-quality teaching and learning in the discipline.	The assessment 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. Assessments should attend carefully to the learning of teachers and students alike and are designed such that teachers also feel they learned something meaningful about their practice. What assessments signal, measure, and provide information about should directly speak to the actions and decisions we want students, educators, and leaders to make—and help them learn 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).

Table 2 (continued)

Principle. Instructionally relevant assessment systems are intentionally designed to be…	Summary Statement
Reflective of and Responsive to Learners. Assessments follow principles of universal design and cultural responsiveness to ensure 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 given/current cohort of students. However, it should be noted that timing is not necessarily a discerning feature—more timely assessment results are only useful to instruction 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 to primarily 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 in order 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: Badrinarayan (2025)

Below, each design principle is further developed with rationale, critical features, and examples of systems as appropriate.

Principle 1: Authentic. Assessments should highlight and center the key concepts, modes of inquiry, and ways of learning in the discipline.

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 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 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 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.

In practice, instructionally relevant assessments that prioritize authenticity include the following essential features:

1. **Include high-quality performance tasks that mirror real world disciplinary activities that require the targeted domain.** Assessments should leverage high-quality performance tasks that mirror (in developmentally appropriate ways) how the key ideas and practices of the discipline are leveraged in real world contexts as a component of state assessment systems. Doing so may involve (1) centering tasks on robust phenomena, problems, and scenarios that ask students to apply multiple disciplinary practices, crosscutting concepts, and conceptual ideas (potentially across disciplines); (2) collaborative investigations, modeling, and argumentation; (3) student-driven research, synthesis, and communication of findings; and (4) student thinking in response to phenomena, problems, and scenarios or contexts that contain authentic uncertainty, requiring students to use their knowledge and practice to address.
2. **Focus on the most-needed instructional shifts and higher-leverage aspects of standards and other related learning goals.** All large-scale assessments are an act of sampling. Federal law in the United States requires that states assess the depth and breadth of their state standards (U.S. Congress, 2015)—while this is intended to ensure that decisions about resource allocation and consequences for school systems are based on comprehensive information, efforts to meet this requirement often obscure a truth about state assessments: that they simply cannot measure everything students need to know and be able to do described by state standards. Current approaches to state test design almost universally value breadth and coverage, giving tests a façade of "measuring everything"—but in doing so, they are most often *not* measuring depth, robust and iterative meaning-making, deeper engagement with disciplinary practices, etc. Instead, assessment design could make a different tradeoff. In a world where no test is actually measuring everything, states and assessment developers should carefully consider *the highest-priority instructional shifts state assessments can support and ensure those same elements are prioritized in the assessment design.* This includes both major instructional shifts (e.g., toward integration of practices and discourse with content) as well as elements of the standards and related learning goals that are going to (a) have the biggest impact on instruction, and (b) be most critical for future learning, when prioritizing how learning goals are translated to assessments.

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 the science and engineering practices, particularly around interpreting and using evidence and being able to be critical consumers of information (National Research Council 2012, NGSS Lead States 2013, Badrinarayan 2024). Accordingly, states may design assessments that foreground science and engineering practices, emphasizing using science ideas as part of making sense of provided data and distinguishing between the validity of different claims and sources (National Assessment Governing Board 2024). 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—in practice, this would likely look like having fewer score points associated with each disciplinary core idea, and more score points associated with integrated bundles of practices and core ideas, connected to more integrated performances. Similarly, across math, English language arts, and science, two major shifts researchers and practitioners emphasize is the need for students to (1) incorporate discourse and social meaning-making into learning activities, and (2) integrate the knowledge and skills across many different standards. Assessments could be designed to both signal these kinds of shifts: for example, large-scale on-demand assessments could develop 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. Large-scale assessments could also leverage emerging technologies to simulate discourse between students and interactive agents, or leverage collaborative and discourse-centered activities from the classroom in on-demand assessment activities (Shepard 2000).

Principle 2: Curriculum-Anchored. Assessments are connected to, and informed by, high-quality curriculum.

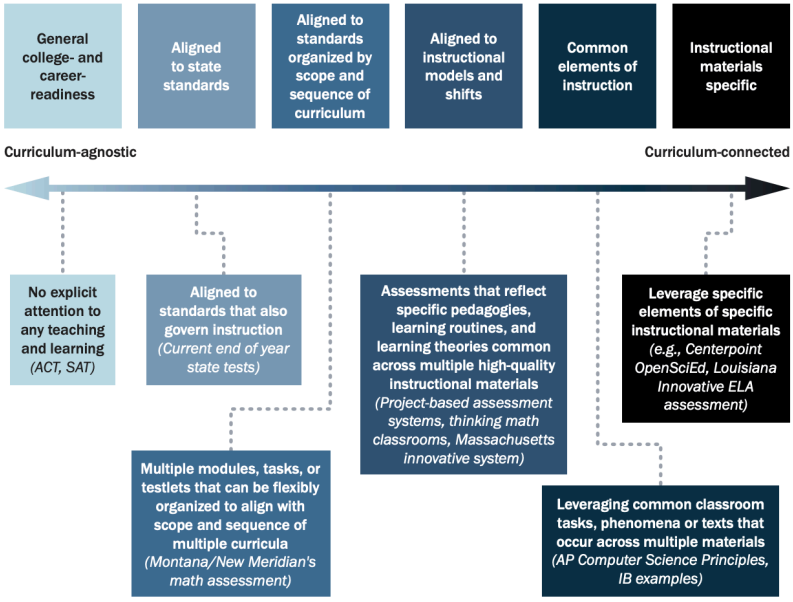
Traditional approaches to assessment design assume a unidirectional flow of influence in which state standards drive curriculum and assessment design independently, and, by virtue of alignment to standards, the assessments and curriculum are aligned. This "transitive property of alignment" approach makes 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 (National Academies of Sciences, Engineering, and Medicine, 2018). As a result, "curriculum agnostic" 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 (Dadey & Badrinarayan 2022, Badrinarayan & Steiner 2023).

Instead, states should consider designing assessments that intentionally leverage features of high-quality curriculum (Dadey and Badrinarayan 2022, Badrinarayan and Steiner 2023). Across disciplines, leaders and developers have taken great care to ensure that high-quality instructional materials[1] (HQIM) and associated professional learning reflect the best evidence we have on how students learn within each of the disciplines, and do so in service of building mastery toward commonly established standards and related learning goals (e.g., EdReports n.d.). 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).

1 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

Figure 1
Ways to Anchor Assessment to Curriculum

Source: Adapted from Badrinarayan 2024.



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 (a) incentivize the use of high-quality curriculum approaches and materials by making clear these approaches will help students be successful on state assessments, and (b) 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 this design principle assumes 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 (Ruiz-Primo et al., 2012, D'Agostino et al., 2007). 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.

Instructionally relevant large-scale assessments intentionally account for curriculum by:

1. **Anchoring to curriculum in ways that reflect intended and likely use.**
   There are many different ways to make meaningful connections between high-quality curriculum and a state's assessment system (See Figure 1). The most appropriate connection within a given context 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, it will be important for states to 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 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. **Providing explicit guidance for how the assessment approach can be used with multiple HQIM 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. 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 are assessing 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; 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. **Offering 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 there are systemic opportunities for educators to develop an understanding of how to design and use assessments to support their specific

curriculum contexts. This aspect builds on evidence suggesting curriculum-based professional learning is particularly impactful on building instructional practice for teachers and meaningful learning for students (Short & Hirsch 2022).

Such a system 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.

Principle 3: Educative. Assessments build educator understanding of effective teaching and how students learn in the discipline.

A major concern about current assessment systems is they are used to justify ineffective, sometimes harmful, teaching and learning practices (NEA, 2021a, 2021b; AFT, 2024; Armstrong, 2013). 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 (a) build educator understanding of what high-quality teaching, learning, assessment, and feedback cycles in science look like, and (b) 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 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 by the very fact of their implementation. 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 assessment scores alone often drive short-term moves that do little to support student learning and retention (e.g., adding a unit on the scientific method to address perceived student deficiencies in science and engineering practice, repeatedly practicing reading decontextualized reading 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 their discussions shift toward deeper understanding of state standards, needed shifts in practice, and concrete next steps.

This principle 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. Systems that are designed to be educative may:

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 (a) make more principled decisions about how to interpret and use reported data, (b) understand what it looks like for students to demonstrate targeted learning in context, (c) connect student performance back to their own teaching practice, and (d) notice nuances in student thinking and performance to help pinpoint needed support for individual and groups of learners.

2. Center and value 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 reflection 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 assessment tasks follow a coherent storyline from the student 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 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, assessments that illustrate high-quality teaching 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, assessments could provide detailed annotations that focus on associated teaching and learning practices (as opposed to only highlighting alignment) with supporting resources linked.

3. **Ensure 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 stakeholder (rather than generic "scientist" or "students");
   - 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 science 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 provision 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.

Involve educators in assessment development, evaluation, and scoring efforts. As discussed above, 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.

Principle 4: Developmental and Asset-Oriented. Assessments recognize what students do know and can do, and surface progress relative to students' own performance along appropriate learning progressions.

Educators often note 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, 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 (O'Donnell & Sireci 2021; Papay et al., 2011; Boaler, 2011; Carey, 2014).

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 students learn by building on prior understanding and experience, rather than simply accumulating isolated pieces of knowledge and skill. Assessments designed to surface facets of student understanding along within-year and multiyear learning progressions for all

three dimensions of science standards can help students, families, educators, and leaders better identify assets of student thinking that can be leveraged as part of future learning opportunities.

Developmental and asset-oriented large-scale assessment systems may:

1. **Assess learning along learning progressions.** Instructionally relevant assessments should allow educators and students 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 strategy is essential for students who may not have had access to high-quality disciplinary instruction to still demonstrate assets in thinking that can be leveraged for next steps.

3. **Ensure 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 on 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 service of 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 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.

Principles 5: Reflective of, and responsive to, learners. Assessments follow principles of universal design and cultural responsiveness to ensure each learner is supported in making their thinking visible.

All students have and are part of rich cultural traditions that govern how they learn, understand, and make thinking visible (Nasir et al., 2020, National Academies of Science, Engineering, and Medicine, 2018). Traditional approaches to state assessments that focus on standardization and "neutral" assessment contexts do not acknowledge 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. Thus, many current state assessments are not measuring what they intend to, inadvertently reporting on students' language and cultural familiarity rather than their science understanding and practice. 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 (Randall et al., 2021, Randall et al., 2022).

Assessments that are reflective of and responsive to learners may:

1. **Assessments should act as windows and mirrors for students. Student performance on assessments are 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, and the goal of assessment design is to vary this experience such that (a) all students feel seen and represented within an assessment, and (b) 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 representation 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.
- Position assessment as learning opportunities—students should feel like 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 (a) the phenomenon/problem-based scenario and contextual information provided, (b) the items and student responses, and (c) reporting and feedback processes. It is essential these connections be as authentic as possible to avoid well-intentioned essentializing, stereotyping, and inappropriate assumptions about what is important to students and communities. This authenticity may look like:
   - leveraging known experiences from curriculum and high-quality instructional materials (connecting to Principle 2);
   - 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 "students" and "communities" are not static, monolithic entities—both communities and students are dynamic and ever evolving. This note suggests 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 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 (a) more than stereotyped or victim-centered experiences, and (b) 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 (a) reflect a range of disciplinary ways of knowing (e.g., that there are multiple ways to "do science"); (b) 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 (c) 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.

Principle 6: Useful for informing decisions that impact instruction. Assessments are designed to produce relevant information at appropriate times to support decision-making.

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. 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 makes it clear that across stakeholder 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 stakeholders with the right information at the right time to drive positive instructional shifts in the discipline?" It should be noted that attention to likely assessment use must be part of the assessment design and development process for instructionally relevant assessments. Put another way, it is not enough for assessments to simply make information about student progress or proficiency available, assuming that any decisions about what to do based the information is the responsibility of other parts of the education systems within which the data are used—instead, assessments intended to be instructionally relevant take seriously the impact of assessments (consequential validity) and design accordingly (Messick 1989, Iliescu and Greiff 2021).

Assessments that meaningfully attend to use may:

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. Consideration of claims 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 "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. Doing so 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.
   - *Focusing on a broader range of information* that goes beyond scores and sub-scores, 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 such that their content and timing aligns with intended use, and designs for likely use.** While it may seem like more frequent information (e.g., via through-year assessments administered three or more times a 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 it is imperative assessments designed to be administered more frequently are either (1) worth the disruption and potential costs to teaching and learning, or (2) designed to become a value-added component *of* instruction. 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.(Goertz et al., 2010, Dadey et al., 2023, Datnow and Hubbard 2015) 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 influence 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.

Conclusion: Making a Different Set of Trade-offs

When state education departments view instructional impact as just as important as producing data to serve program monitoring, they can make different trade-offs in system design (See Figure 2), ultimately yielding better information and better educational experiences for young people.

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 for better assessments.

Figure 2

	Perspective A: the primary focus of state assessment design should be on program monitoring. Positive instructional impact is nice to have, but not a must have.
Information prioritized	• Answering the question: are students proficient in the standards? • Coverage of as many standards as possible • Aggregate scores that can be used for comparing overall performance for schools, districts, and student subgroups at the macro level • Achievement/performance level descriptors relative to grade-level standards
Funding Allocation	• Assessment vendors and technical assistance partners who support test development, psychometrics, test maintenance, etc.
Scoring	• Prioritize machine scorable items. • Teacher involvement is a nice to have, not a must-have; teachers may be invited to participate in some content development and review activities, but teacher involvement is not central to the goals of the assessment. .
Student experience	Students experience the test as completely separate from instruction; see it as an external judgement of their abilities in the tested area.
Implications for test design	• Prioritize selected response items that can be developed and scored quickly and at low cost. • Prioritize coverage, focus on the of the easiest-to-assess content • Acceptable trade-off that teachers, students, families do not see the assessment items, and only see score reports

Figure 2 (continued)

	Perspective B: the primary focus of state assessment design should be on positive instructional impact. The assessment should also surface program monitoring information, but positive impact on instruction is a non-negotiable.
Information prioritized	• Answering the question: how well do students understand and use the standards? • Sufficient but not comprehensive coverage, emphasis on opportunities to measure depth, disciplinary practices, and deeper learning • Individual and aggregate scores as well as examples of tasks and student work that help educators and leaders make shifts that improve instruction • Achievement/performance level descriptors that more precisely pinpoint what students know and can do (e.g., along multiyear learning progressions).
Funding Allocation	• Assessment vendors and technical assistance partners • teachers, LEA, and SEA capacity • Substantial proportions of the resources are more directly supporting teaching and learning rather than only test-related activities.
Scoring	• Prioritize teacher-scored tasks alongside machine-scored items. • Teachers are an integral part of assessment development and scoring because participation in these activities builds educator understanding of standards and the kinds of teaching and learning that are needed to excel in the discipline (and on the assessment).
Student experience	Students experience the assessment as coherent with their instructional experiences, and as a meaningful experience in its own right. The assessment is engaging, motivating, and connected to what students have been asked to engage with in the classroom.
Implications for test design	• Include tasks that model instructional shifts and sophisticated performance expectations. • Prioritize surfacing information that signals and provides information about the most important instructional shifts, including disciplinary practices, disciplinary sense-making, and application of concepts and practices to relevant situations. • Teachers must have access to at least a subset of the assessment tasks to build practice and capacity.

References

American Federation of Teachers Assessment Task Force. (2024). *Real solutions for improving assessment.*

Armstrong, T. (2013). *15 reasons why standardized tests are problematic.* ASCD.

Badrinarayan, A. (in press). *Instructionally Relevant Assessments: What is the Role of Performance Assessments.* Learning Policy Institute.

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

Badrinarayan, A., & Steiner, D. S. (2023, June). *Positioning state assessment systems in service to teaching and learning: The role of high-quality curriculum in state assessment design.* Education First. https://www.education-first.com/wp-content/uploads/2023/06/positioning-state-assessment-systems-in-service-to-teaching-and-learning.pdf

Bennett, R. E., Darling-Hammond, L., & Badrinarayan, A. (Eds.). (2025). *Socioculturally responsive assessment: Implications for theory, measurement, and systems-level policy.* Routledge.

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.

Dadey, N., & Badrinarayan, A. (2022, April 21). *In search of the 'just right' connection between curriculum and assessment.* Center for Assessment. https://www.nciea.org/blog/in-search-of-the-just-right-connection-between-curriculum-and-assessment/

Dadey, N., Evans, C., & Lorie, W. (2023). *Through-year assessments: Ten key considerations.* National Center for the Improvement of Educational Assessment.

Darling-Hammond, L. (2017). *Developing and Measuring Higher Order Skills: Models for State Performance Assessment Systems.* Learning Policy Institute and Council of Chief State School Officers.

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

Datnow, A., & Hubbard, L. (2015). Teachers' use of assessment data to inform instruction: Lessons from the past and prospects for the future. *Teachers College Record, 117*(4), 1–26. https://eric.ed.gov/?id=EJ1056748

EdReports. (n.d.). *Our process.* EdReports. https://edreports.org/process

Evans, C. M., & Taylor, C. S. (Eds.). (2025). *Culturally responsive assessment in classrooms and large-scale contexts: Theory, research, and practice.* Routledge.

Goertz, M. E., Oláh, L. N., & Riggan, M. (2010). *From testing to teaching: The use of interim assessments in classroom instruction (CPRE Research Rep. #RR-65).* Consortium for Policy Research in Education. https://repository.upenn.edu/cgi/viewcontent.cgi?article=1023&context=cpre_researchreports

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

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 Assessment Governing Board. (2023). *2028 NAEP science assessment framework.* https://www.nagb.gov/content/dam/nagb/en/documents/publications/frameworks/science/2028-naep-science-framework.pdf

National Education Association. (2021a). *Beyond the bubble: Americans want change on high stakes assessments.*

National Education Association. (2021b). *Principles for the future of assessment.*

National Research Council. (2012). *A framework for K–12 science education: Practices, crosscutting concepts, and core ideas.* The National Academies Press. https://doi.org/10.17226/13165

NGSS Lead States. (2013). *Next Generation Science Standards: For states, by states.* The National Academies Press. https://doi.org/10.17226/18290

O'Donnell, F., & Sireci, S. G. (2022). *Language matters: Teacher and parent perceptions of achievement labels from educational tests. Educational Assessment, 27*(1), 1–26. https://doi.org/10.1080/10627197.2021.2016388

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.

Randall, J., Poe, M., & Slomp, D. (2021). *Ain't ought to be in the dictionary: Getting to justice by dismantling anti-Black literacy assessment practices. Journal of Adolescent & Adult Literacy, 64*(5), 594–599.

Randall, J., Slomp, D., Poe, M., & Oliveri, M. (2022). *Disrupting White Supremacy in Assessment: Toward a Justice-Oriented, Anti-Racist Validity Framework. Educational Assessment, 27*(2), 170–178.