



Performance Assessments: How State Policy Can Advance Assessments for 21st Century Learning

**By Ace Parsi and
Linda Darling-Hammond**

White Paper, January 2015

A draft of this paper was released at the NASBE October 2014 Annual Conference.

Abstract:

Employers, postsecondary institutions, and civic leaders are urging greater focus on 21st century skills essential for college, career, and civic success: problem solving, interpersonal skills, and collaboration, among others. In response to these demands, states across the country are working to realign policies—on learning standards, assessments, and human capital strategies—to set a new course for their state education systems.

Many states are reconsidering their assessment strategies and asking whether existing assessments are designed to be the barometer, resource, and engine for learning necessary to support the new demands on students. To address gaps in existing assessments, many states are considering the role of performance assessments—assessments that require students to craft solutions to problems by constructing an answer, producing a product, or performing an activity rather than selecting from a list of multiple-choice answers—within their statewide assessment strategies. By requiring students to complete multifaceted tasks, these assessments can better support 21st century learning.

This report is intended to familiarize state boards of education with performance assessments and help state board members and other policymakers address some of the thorniest issues around these assessments: purpose, sustainability, reliability, accountability, policy alignment, equity, professional practice, and implementation. The report ends with a set of discussion questions so that each state can begin to analyze barriers and opportunities toward effective implementation of these assessments.

NASBE's deeper learning work is funded by the William and Flora Hewlett Foundation. NASBE thanks the foundation for its support. The views expressed in this white paper, funded in part by Hewlett, are those of the authors alone.



Barometer-Assessments of learning can show how students are progressing in meeting state standards and developing knowledge, skills, and dispositions deemed important for college, career, and civic success. In the process, they can help direct needed resources and supports to district, schools, and students.



Resource-Assessments for learning can help inform educator practice by identifying students' learning progress and needs. They can also inform students about how to improve as they revise their work and set goals for developing their skills. They can give parents insights about how to support their children's learning.



Engine-Assessments as learning offer meaningful tasks that support and deepen learning as students engage in them. They offer teachers good models of instruction.

Performance Assessments: How State Policy Can Advance Assessments for 21st Century Learning

Assessments *of*, *as*, and *for* learning play essential roles in advancing learning outcomes: as a barometer of student learning, as a resource for educators, and as an engine driving learning. Assessments are not in themselves ends—they are means to support student learning. This paper asks two essential questions regarding whether assessments are achieving their purposes. First, are the assessments that states are using aligned to the knowledge, skills, and dispositions deemed essential to students' future success? If not, what is needed to better align the state's primary goal—college, career, and civic readiness for all students—and state assessment policy?¹

At present, there is a mismatch between the knowledge, skills, and dispositions states expect students in the 21st century to attain upon high school graduation and the skills most current summative assessments measure and support. Many states have not yet begun to create new assessment systems to evaluate these more complex skills. Such systems will need to place greater emphasis on performance assessments—that is, assessments that require students to craft their own responses to problems through constructing an answer, producing a product, or performing an activity rather than merely selecting from multiple-choice answers.

Performance Assessment (active noun)-
Assessments that require students to craft their own responses to problems through constructing an answer, producing a product, or performing an activity rather than merely selecting from multiple-choice answers.

The purpose of this paper is to help state board of education members and other policy leaders understand the contributions that performance assessments can make as part of a comprehensive assessment and accountability system and to outline key considerations that will enable states to use such assessments to advance their college, career, and civic readiness goals. Because states have different histories and experiences, the paper includes a set of discussion questions that will enable state leaders to assess their capacity to use and advance performance assessments.

Educational Imperatives and the State of Assessments

For many educators and state leaders, performance assessments are not new. —In fact, a number of states had made significant headway in adjusting their assessment strategies prior to the inception of the No Child Left Behind Act of 2001, otherwise known as NCLB. What early adopter states such as California, Washington, Kentucky, Vermont, and Maine knew then—and is even more evident today—is that multiple-choice tests fail to measure higher order skills. The need to support students and educators in meeting higher-order learning goals creates a compelling case for state investments in performance assessments. These assessments should meet three educational imperatives: excellence, equity, and instructional improvement.

Excellence. For decades, business surveys and labor market reports have highlighted that, as routine tasks have become automated, the labor market has shifted from requiring routine skills to a greater emphasis on interactive and analytic skills and tasks that place higher cognitive demands on the workforce.² In 1970 the skills most valued by Fortune 500 companies were reading, writing, and arithmetic. At the start of the 21st century, the top-valued skills were teamwork, problem solving, and interpersonal skills (see table).³ Surveys of postsecondary faculty and employers confirm that students need to be critical thinkers, creative problem solvers, and effective communicators and collaborators if they are to succeed in college and careers.⁴

Fortune 500 Most Valued Skills

	1970	1999
1	Writing	<i>Teamwork</i>
2	Computational Skills	<i>Problem Solving</i>
3	Reading Skills	<i>Interpersonal Skills</i>
4	Oral Communications	Oral Communications
5	Listening Skills	Listening Skills
6	Personal Career Development	Personal Career Development
7	Creative Thinking	Creative Thinking
8	Leadership	Leadership
9	Goal Setting / Motivation	Goal Setting / Motivation
10	<i>Teamwork</i>	Writing

Unfortunately, too many states' assessments fall far short of this new excellence imperative.

Analysis by the RAND Corporation found that only 2 percent of mathematics items and only 20 percent of English language arts items on current state tests measure higher order skills.⁵ While the Partnership for Assessment of Readiness for College and Career (PARCC) and Smarter Balanced Assessments promise higher levels of cognitive rigor, they do not address skills such as the ability to collaborate, communicate, conduct research, use technologies, and plan, persevere, and overcome obstacles during complex tasks.⁶

Equity. The equity imperative has deep roots within the collective national psyche—namely, that anyone, regardless of the circumstances into which they were born, should be able to realize their fullest potential through individual effort and merit.⁷ The equity imperative is intertwined with the excellence imperative: If *all* students are to succeed, then *all* students must be provided a rich and meaningful education that prepares them for 21st century success.

The NCLB Act set ambitious goals for driving educational systems toward equal opportunity, aiming for all students to be proficient in reading and mathematics by 2014. While the overall equity goal was significant and positively contributed to the national conversation, when

implemented, the law's narrow provisions around testing had the unintended consequence of exacerbating inequality: Schools serving low-income students and students of color, for example, began to alter their educational approach to focus almost exclusively on state tests, which have been dominated by less cognitively demanding formats.⁸

Instructional Improvement. Neither excellence nor equity can be fully achieved unless states, school districts, and their schools leverage assessments toward giving educators more actionable information about student learning. Well-designed assessments share several key characteristics—instructional usefulness, validity, reliability, and sensitivity to the effects of high quality, cognitively rich instruction.⁹

Current tests offer limited support for teacher learning and instructional improvement. Because they do not reflect and embody the skills students need, they cannot inform instruction that aims to fulfill these purposes.¹⁰ Additionally, multiple-choice tests do not offer any information about students' thinking or reasoning, their misconceptions, or ability to express their ideas. Thus they provide little information to help teachers shape instruction. Last but not least, since these assessments are primarily summative, they cannot inform continuous instructional improvement, a goal facilitated by a rich array of assessments within a comprehensive state assessment system.¹¹

In contrast to multiple choice summative assessments most states emphasize, performance assessments are naturally geared toward promoting excellence, equity, and instructional improvement. When performance assessments call on students to synthesize knowledge and to design and create solutions to complex problems, the learning that results demonstrates higher cognitive complexity. When performance assessments require students to complete multifaceted tasks that involve finding resources, analyzing and synthesizing information, getting and using feedback, they are also promoting “soft” skills like planning, perseverance, grit, and resourcefulness, which are important to students' college, career, and civic success. And when educators use the results of such assessments to inform and revise their curricula and instructional practices, they in turn strengthen an education system that is better aligned to excellence for every student in the system, regardless of their circumstances.

How Might Performance Assessments Be Used in a State System?

The discussion on how to make assessments effective will not be new to most states. Even better, many of them had already begun to make significant strides toward performance assessments before the inception of NCLB, which increased the number of mandated state tests, changed the required nature of those tests, and thus led states to scale back on their efforts to change assessment. According to a study by the US Government Accountability Office, state use of multiple-choice tests dramatically increased following the law's passage.¹²

At the same time, school districts in Australia, Hong Kong, and Singapore began to expand the use of performance assessments.¹³ Often, standardized tests and classroom-embedded

performance assessments are combined. In Queensland, Australia, for example, national tests of numeracy and literacy occur in grades 3, 5, 7, and 9. In other grades, common statewide performance tasks are tested. An example of a common task in grade 7 science and one expected of students at the senior level are shown in the boxes below. Clearly, these Australian students are expected to develop sophisticated, college-ready skills, and this learning is supported by the assessment strategy.

QCAT for 7th grade science:

90 minutes over 1-2 days;
Given some contextual information, students must analyze and construct food webs in two environments. Through multiple prompts, students must show an understanding of food chains and the impact of environmental disruptions on populations.

Extended Experimental Investigation at the

Senior Level

Over four or more weeks, students must develop and conduct an extended experimental investigation to investigate a hypothesis or to answer a practical research question. Experiments may be laboratory or field based. The outcome of the investigation is a written scientific report of 1,500 to 2,000 words. The student must

- develop a planned course of action
- clearly articulate the research question and provide a statement of purpose for the investigation
- provide descriptions of the experiment
- show evidence of student design
- provide evidence of primary and secondary data collection and selection
- execute the experiment(s)
- analyze data
- discuss the outcomes of the experiment
- evaluate and justify conclusion(s)

Local performance assessments are used in every grade. These include extended papers, research projects, analyses, and problem solutions in each subject. They are developed at the school level but are subject to quality control and review of scoring by a state panel. At the high school level, each student's work is collected into a portfolio that is scored using common criteria. Each school's scoring is audited by an external panel of educators and professors from the higher education system.¹⁴ All of this activity is aligned to the state's content standards, called Essential Learnings, which are published in a framework, along with unit templates and guidance for assessments in each subject.

A similar system was used in *Kentucky's* Instructional Results and Information System (KIRIS) between 1992 and 1998. KIRIS included standardized tests in grades 4, 8, and 11, featuring multiple-choice and short-essay questions supplemented by performance tasks in mathematics, which required students to solve applied problems, and writing portfolios, which collected specific samples of classroom work over time. The writing portfolio continued until 2012 in grades 4, 7, and 12, while an On-Demand Writing Assessment was used in grades 5, 8, and 12. The state trained teachers to score the portfolios at the school level using analytical rubrics and processes that assured validity and reliability: two readers, using double-blind rubrics, were charged with scoring each piece in students' portfolios. The system became so successful that by 2008 the agreement rate among readers who audited school-level scores was over 90 percent, and teachers and administrators alike saw the assessment as a valuable tool to inform meaningful student learning.¹⁵

New Hampshire is following a similar path. In addition to developing new summative assessments through the Smarter Balanced Assessment Consortium, New Hampshire's assessment system will include common statewide performance tasks with high technical reliability in core subjects, valid and reliable locally designed assessments, a web-based bank of common performance tasks, and a network of practitioner "assessment experts" to support schools. This assessment strategy couples high expectations embedded within statewide college- and career-ready competencies, common statewide performance tasks, local autonomy, support delivered through professional development institutes and regional networks, and district peer review audits. New Hampshire hopes that its efforts "promote the use of authentic, inquiry-based instruction, complex thinking, and application of learning ... [and] incentivize the type of instruction and assessment that support student learning of rich knowledge and skills."

In these and other states' efforts, success in the support and implementation of performance assessments reflects common themes:

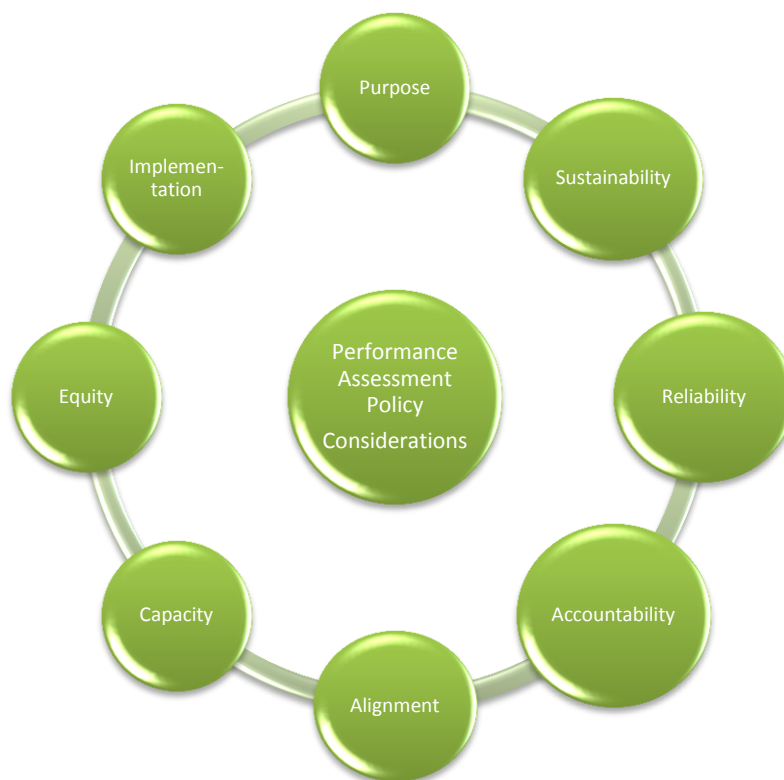
- a commitment to knowledge, skills, and dispositions deemed important for all students' future success that represents the foundation of the overall state assessment strategy;
- the capacity to incorporate performance assessments within a comprehensive state assessment and accountability system;
- a concerted effort to ensure the state's educators are substantially included in the development, implementation, and scoring of these assessments; and

- a long-term commitment to promote continuous systemic improvement, while capitalizing on efficiencies such as consortium efforts where possible, facilitating collaboration and pooling of resources to achieve better results.

Each of these state efforts required thoughtful and concerted collaboration of policymakers and practitioners to ensure that the whole assessment and accountability system served intended learning purposes. To that end, we include some key questions states must answer to facilitate development of such a system.

Cracking the Performance Assessment Nut: Key Questions for State Board of Education Members

States face a common set of issues in creating systems of assessment that support excellence, equity, and instructional improvement. This section addresses some of the thorniest: purpose, sustainability, reliability, accountability, policy alignment, capacity, equity, and implementation. The worksheet questions that accompany this document serve as an additional resource to help state board of education members do a more targeted gap analysis of strengths and challenges within their states and assess steps they can take to advance performance assessments within unique contexts.

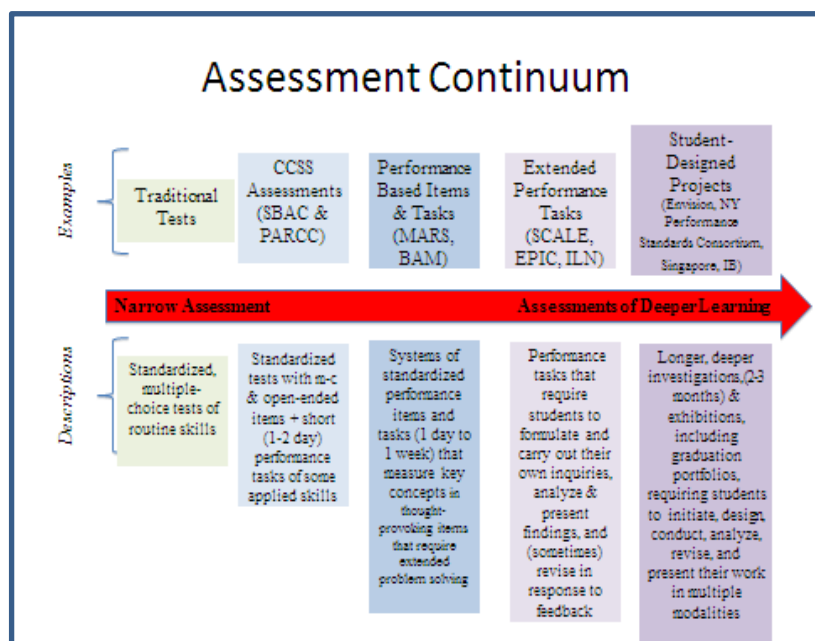


Purpose: What is being assessed and to what end?

Assessments serve a variety of important functions, including as barometer, resource, and engine of learning. In order to serve these functions effectively, policymakers must ensure every assessment serves a clear purpose and fits within a broader state and local assessment strategy designed to effectively

support learning. Unless states carefully figure out how to do this, students and educators could feel overtested and underinformed, leading to lost instructional time and frustration across the system, a frequent criticism of many narrow summative assessments in place across the country.¹⁶

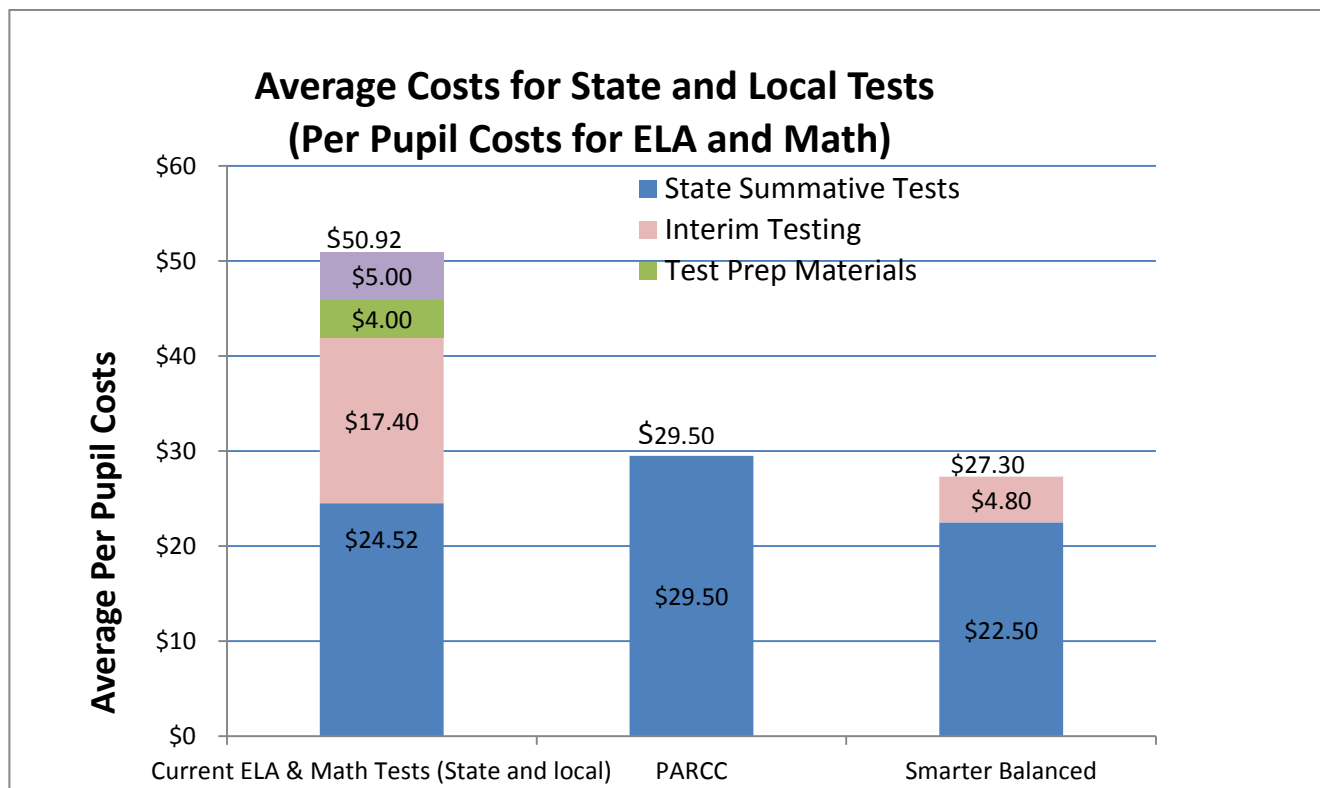
A comprehensive system that delineates the purpose of each assessment can include instruments across a continuum of assessments (see figure). In order to evaluate skills essential to college, career, and civic success, some of the assessments should fall toward the right side of the continuum and be used as periodic common measures, elements of end-of-course examinations, assessments for competency determinations in competency-based systems, or elements of graduation portfolios. There can be fewer, higher quality state tests if they are used as a barometer to validate the results of richer assessments administered locally rather than to serve all purposes.



Sustainability: How can the state pay for new performance assessments?

Like most good things, a fair, valid, and reliable assessment and accountability system comes with a price tag, which is an immediate concern for states struggling to raise education spending back to pre-recession levels. According to an analysis by the Center for Budget Policies and Priorities, as of the 2013–14 school year, 35 states were providing less funding per student than they were prior to the 2007–2009 recession.¹⁷ Despite such shortfalls, state policymakers should prioritize investments in higher-quality assessments for one simple reason: States currently spend only about a quarter of 1 percent of total K-12 education expenditures (about \$25 per pupil for NCLB required reading and math tests) on assessments that have been found to be of relatively low quality, yet they base many decisions on these tests, thus focusing schools’ efforts almost exclusively on low-level skills. Ironically, local districts spend even more on test preparation, plus interim and benchmark tests aimed at improving performance on the state tests, bringing total spending to over \$50 per pupil. This spending does not improve higher-order learning, however, because nearly all of it is geared toward boosting scores on multiple-choice tests.

By comparison, consortia tests cost between \$25 and \$30 per pupil for formative and interim assessment tools, as well as summative tests in English-language arts and mathematics, a small amount compared with other educational investments (see figure). (Consider, for example, that a single textbook can cost \$100 or more.) Costs are reduced by participating in state consortia, by strategic uses of technology, and by thoughtful use of teacher-moderated scoring, which can also have added professional learning benefits. A new consortium, the Innovation Lab Network sponsored by the Council for Chief State School Officers, is also creating a bank of validated performance assessment tasks, rubrics, and scoring protocols that states can tap to create broader assessment systems. By spending smarter and aligning initiatives, states could in the long term not only reap the benefits of performance assessments but spend less.



Reliability: How can a state know assessment results can be trusted?

Early efforts at performance assessment in the 1990s, such as Vermont’s portfolio system, raised concerns about inter-rater reliability in scoring. Since then, scoring reliability has been achieved when tasks are designed with common templates and specifications, when teachers with training use common rubrics to do the scoring, and when teachers are engaged in a moderation or auditing process to achieve consistency. As noted earlier, Kentucky’s writing portfolio reached a high level of scoring reliability when these features were put in place, as did performance tasks in a number of other states. Performance assessments in systems like the Advanced Placement and International Baccalaureate programs, and like those used overseas, have long been reliably scored in these ways. In formative contexts, teachers typically score their own students’ work; in summative contexts, teachers typically score the work of students from other classrooms or schools. Either way, there are added benefits to bringing teachers together to engage in training and moderation processes that help them score performance assessments reliably: When teachers come together to look at student work, discuss the standards and criteria used to evaluate it, and reflect on

student learning and teaching, they become more knowledgeable about standards and how to teach them. This process of creating clear, common expectations thus not only improves instruction, it is also essential to ensuring assessments results have validity for a variety of external stakeholders, including postsecondary institutions, and provide useful information to educators as students move from school to school.

Capacity: What will this mean for educators?

A major reason to use performance assessments is because they help educators identify student needs and can be a launching point for cognitively rich learning experiences. But performance assessments can promote such benefits only when educators are given the necessary time, training, and support to achieve the level of ownership and investment realized in states such as Kentucky, Vermont, and Maryland.

Successful states help teachers learn to use performance assessments in the classroom; they also incorporate educators in the development of rubrics and performance tasks as well as scoring tasks—both for their own students and for students elsewhere in the state, depending on the purpose of the assessments. These states also train educators on how best to use performance assessment results to further student learning. States have found that when both assessment and reflection on assessments are more integrated into an educator’s day to day practice, the result is a system that can better leverage the full benefits assessments provide.

Alignment: How does this fit within the rest of the state’s policies?

When states integrate assessments with professional learning they highlight a key fact: No one policy exists within a silo, and nearly all are interrelated. When state actions around assessments, accountability, professional learning, use of time, funding, and other policies reflect different priorities and are not mutually reinforcing, negative consequences ensue, including a compliance mentality at the state and local level, competing priorities, and wasted resources.¹⁸ Thus a new assessment strategy could lead to educator innovation fatigue rather than excellence, as educators decide to ride out new policies under the guise that “this too shall pass.”

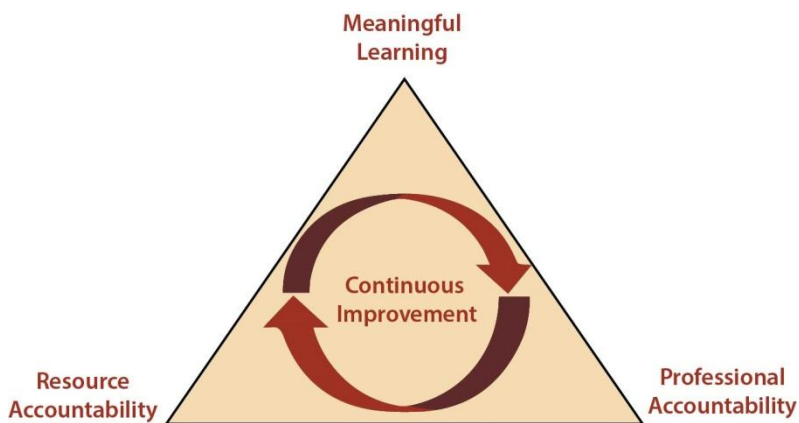
With this in mind, state board of education members and other policy leaders should analyze existing state policies and reflect on which will influence the overall success of new assessments before taking up performance assessments policy. Supportive policies should include considerations around aligned educator preparation, ongoing learning, and evaluation systems, statewide accountability systems, and policies on school finance, calendar, and credit policies. By addressing these and other key considerations, state leaders can ensure that assessment policy is a key part of a comprehensive statewide strategy to promote college, career, and civic success.

Accountability: How should assessment results be used?

Achieving policy alignment requires clarity on the role each policy plays within a comprehensive statewide educational strategy. Clarity is especially needed around accountability and assessment policy, two important policy areas that have become synonymous in too many conversations but in reality should

remain distinct. While assessments can be part of a comprehensive accountability system—providing useful information to guide investments, curriculum decisions, and professional learning, for example—testing does not create accountability. No test—performance assessments included—should be used as the sole or primary basis of consequential decisions for students, teachers, or schools. Giving tests such a large role in accountability not only overtaxes their limits, it dilutes the other essential functions of assessments: informing educator practice and individual student learning goals.

In other words, assessment data should be used, along with other relevant information, to inform change and drive continuous improvement, not to mete out sanctions and punishments. For performance assessment systems to achieve maximum effectiveness, they must be nested within a broader, integrated system that encourages and reflects **meaningful learning**, provides **resource accountability** that ensures



equitable access to necessary learning opportunities, and offers **professional capacity building** to ensure educators can teach the standards effectively to diverse learners.¹⁹

Many states are developing multiple measures systems for accountability reporting that will facilitate a broader, more intelligent basis for accountability actions. A dashboard of information can include a range of assessments (e.g., state test scores including scores on performance tasks or portfolios, English proficiency scores, AP or IB test results); graduation rates; school holding power; school climate evidence; completion of college- and career-ready courses of study; measures of resources; and other data that reveal the quality of education students receive and the outcomes they attain.

Equity: How does the state ensure assessments meet the needs of all students?

Strategic use of performance assessments and multiple-measure accountability systems can support equity, ensure that student progress is evaluated in a more nuanced, comprehensive manner, and ensure that resources are allocated to meet each student’s needs. Performance assessments provide multiple means for students to demonstrate competencies and knowledge than simply having answers be marked right or wrong. Well-designed performance assessments help English-language learners and students with special needs better demonstrate their knowledge.²⁰ In addition, classroom-embedded performance tasks such as science investigations and research projects ensure that all students engage in activities that foster higher-order skills and that their schools have incentives to provide more of this rich educational diet.

Of course, state assessment developers must take steps to ensure performance tasks are fair, gauge what students know and are able to do, and do not reflect language barriers or disabilities that are otherwise unrelated to assessment of learning tasks. These steps include piloting tasks to ensure that they are understandable to different populations of students and carefully developing and reviewing rubrics,

scoring guides, and instructional tools so that they support meaningful assessment for the full range of students.

Implementation: How should performance assessments get rolled out?

The success of any new policy initiative depends on steps the state takes to ensure the policy is effectively introduced and implemented. Performance assessments are no different. In the absence of judicious design and a well-supported transition period, educators, schools, and even state education officials charged with supporting and implementing new performance assessments could feel overwhelmed. States must thus balance the urgency to act with the patience needed to build and support a valid and reliable system and to seek stakeholder investment and public understanding of the usefulness of assessment results.

To help achieve this balance, state boards of education should ensure adoption of new assessments are coupled with a detailed and well-communicated implementation plan, outlining key benchmarks and steps toward success. This plan should allow enough time to engage stakeholders in the design of the performance assessments, solicit input on the development of common performance tasks, promote strategies to ensure the state's educators are effectively trained and ready to use the assessments to achieve better learning results for their students, and explore approaches for reducing costs and aligning assessments with other policies. Through thoughtful implementation and rollout of performance assessments, policy leaders can position their states to fully realize the benefits of these assessments in driving learning and ensure that all students are ready for the world that awaits them beyond the schoolhouse doors.

Conclusion: Cost-Benefit Analysis

Transforming a state assessment and accountability strategy to support and advance the knowledge, skills, and dispositions that are essential to students' college, career, and civic success is not an easy task. States will need to make persistent efforts to finance the system, train educators, and use the information that performance assessments provide to continuously improve state education systems. These costs are dwarfed, however, by the substantial costs of inaction: frustrated educators questioning the value of existing tests, continued and persistent opportunity gaps, and most important, a system that is misaligned to the goal of enabling all students to seize opportunities the 21st century provides.

Luckily, many states have already begun to engage in this important task. The following worksheet questions are designed to serve as a self-assessment and conversation starter to help in identifying potential gaps in conditions that lead to the successful implementation of such assessments and possible alternatives to addressing these gaps. In the end both these essential conversations and their subject, state performance assessments, are necessary interim steps toward the more fundamental goal: high-quality learning that prepares students for college, career, and civic success.

Key Discussion Questions for State Policy Leaders

These questions reflect key issues other successful states and countries have considered when implementing such assessments and do not represent an exhaustive list. Some require discussion and action among board members only. Others will require discussions with other state policymakers, including state education agency staff, governors, and legislatures. Still others will require the board to gather information on the current state of affairs. Because states begin this reflection from a different starting points and with different authorities related to these issues, the answers will also differ—in fact some states will have to do some work prior to even getting to some of these questions. That said, state boards of education and their members are encouraged to reflect on the full range of their influence on these issues—as policy-makers, communicators, and conveners.

Once they have begun the process of building comprehensive assessment systems, state boards should remember that, as with other tasks essential to delivering high-quality education, they will never be done. They should consider answering these questions as an iterative process of learning, acting, evaluating, and revising, all with one ultimate goal in mind: college, career, and civic readiness.

Purpose: What is being assessed and to what end?

- What assessments does the state use now? Which of these assessments are mandated, which are optional, and who is subject to them?
- What are the key knowledge, skills, and dispositions the state has deemed essential to college, career, and civic success?
- Do existing assessments effectively evaluate these knowledge, skills, and dispositions, including higher order skills such as research and inquiry; complex problem solving; applications of knowledge to real-world issues; finding, organizing, and evaluating information; communicating in writing, orally, and using technology; collaborating with others to create a solution or product?
- Which assessments do educators in the state report positively affects and informs their educational practice? What assessment changes, if any, do educators believe would support their practice?

Sustainability: How can the state pay for it?

- In total, how much does the state currently spend on assessments, including costs of the assessments themselves, their administration, scoring, and other assessment-related expenditures such as test preparation materials? What about local expenditures for interim and benchmark tests?
- Which of these costs might be reallocated to support greater use of performance assessments in the state or local assessment system?
- What strategies has the state explored in supporting greater efficiency in overall assessment investments, thus reducing long-term costs?
 - Participation in consortia?
 - Strategic uses of technology to administer and score assessments?
 - Teacher-moderated scoring?
 - Reallocation of existing resources such as professional learning funds that could support teacher scoring and lead to greater synergy between assessment and teacher learning?

- Other actions not listed above?
- How much have or would the above actions reduce overall assessment expenditures?

Reliability: How can a state know assessment results can be trusted?

- Have assessments been field tested and validated by students, educators, outside experts, state postsecondary institutions, employers, parents, community members and other stakeholders?
- Which state-supported assessments are common and standardized across the state and which state or locally supported assessments do districts, schools, and educators have discretion in designing? Do each of these assessments have a commonly understood purpose and interpretation?
- How does the state ensure valid and reliable scoring of assessments?
- If the state does not currently use performance assessments, what additional steps could the state take to introduce such assessments while ensuring validity and reliability of assessment results? For example,
 - Designing common statewide performance assessments?
 - Providing common guides, templates, and rubrics to evaluate assessment results?
 - Facilitating statewide workshops, technical assistance, convenings, and other steps to ensure common understanding of assessments and their purpose?

Capacity: What will this mean for educators?

- Has the state involved educators in the development of performance tasks incorporated into the system of assessments?
- Has the state involved educators in the development of rubrics and instructional tools evaluating student success on performance assessments?
- Are competencies around use of performance assessments incorporated into teacher preparation, licensure, and mentoring programs?
- Has the state, through technical assistance, professional development, and learning standards, supported a system where educators have the capacity to use performance assessments to inform their instructional practice?

Alignment: How does this fit within the rest of the state's policies?

- Which of the policies would the state need to modify to support an effective state assessment system?
 - policies around educator preparation, including standards for teacher and leader preparation, licensure, professional learning, and evaluation;
 - policies around accountability, including high school graduation requirements and school-wide accountability indicators;
 - policies around time, including both the use of teacher time and how time for students is accounted for and credited;
 - policies around school finance and funds schools receive to educate students.

- Which of the policy areas poses the greatest challenge to high quality assessment systems that include performance assessment and how might these policies be modified to support greater overall alignment?
- What is the authority of the board in affecting these policies?
 - *policy authority over an issue*, including those granted by the state’s constitution or legislative order over the whole issue or part of the issue;
 - the *capacity to convene* stakeholders and experts across the state to discuss solutions or draw attention to an issue; or
 - the *ability to ask questions* of the state chief school officer, state department of education, or other stakeholders to draw out nuances of an issue.

Accountability: How should assessments be used?

- What measures does the state use within its current accountability system to evaluate student, educator, school, district, and state success?
- Do the measures highlighted above adequately reflect the learning demands the state requires to ensure all students leave school, college, career, and civic ready?
- What other indicators might be productive to add in creating a multiple-measure accountability system that reflects critical aspects of educational excellence and equity?
- Is the breakdown of existing state assessments—among formative, interim, and summative measures—appropriate in informing meaningful student learning?
- Does the state accountability system reflect multiple functions, including accountability of meaningful learning, accountability of professional capacity to deliver meaningful learning, and accountability of resources to ensure educators and schools are well positioned to deliver on state learning goals?

Equity: How does the state ensure assessments meet the needs of all students?

- What do current indicators reveal about the gaps in access and performance across student groups?
- Has the state conducted a gap analysis to evaluate whether the design of current assessments accommodates the needs of the state’s diverse student population?
- Do all kinds of students have access to a curriculum, with accompanying performance assessments, focused on college- and career-ready skills?
- Have diverse populations—students, parents, and stakeholder groups representing English language learners, students with disabilities, students of color, economically disadvantaged students, geographically diverse students (i.e., rural, urban, and suburban), and other populations—been sufficiently included in considerations around the design and implementation of new assessments?

Implementation: How should performance assessments get rolled out?

- Has the state board of education and state considered a short-, medium-, and long-term implementation plan for new assessments? Does this plan incorporate performance assessments?

- If so, what are the key benchmarks associated with implementation, including necessary stakeholder engagement and investment?
- How will educators and state department of education staff be trained in administering and supporting new assessments?
- What has been the timeline for other successful state education initiatives? What can be learned from these initiatives to influence the rollout of new performance assessment systems?

Notes

¹ Much of the material for this brief is drawn from L. Darling-Hammond and F. Adamson, *Beyond the Bubble Test: How Performance Assessments Support 21st Century Learning* (San Francisco: Jossey-Bass, 2014).

² Frank Levy and Richard J. Murnane, *The New Division of Labor: How Computers Are Creating the Next Job Market* (Princeton University Press, 2004).

³ Russell N. Kolstad and Rosemarie Cassel, “Critical Job-Skill Requirements for 21st Century: Living and Working with People,” *Journal of Instructional Psychology* 25, no. 3 (September 1998): 176, <http://search.ebscohost.com/login.aspx?direct=true&db=f5h&AN=1075972&site=ehost-live>.

⁴ American Management Association, *AMA 2010 Critical Skills Survey* (Washington, DC: AMA, 2010); Jill Casner-Lotto and Linda Barrington, *Are They Really Ready to Work?* (New York: Conference Board, 2006); David Conley, *Understanding University Success* (Eugene, OR: Center for Education Policy and Research, 2003); David Conley, *Redefining College Readiness* (Eugene, OR: Center for Education Policy and Research, 2007).

⁵ K. Yuan and V. Le, *Estimating the Percentage of Students Who Were Tested on Cognitively Demanding Items through the State Achievement Tests*, (Santa Monica, CA: RAND Corporation, 2012).

⁶ Herman and Robert Linn, “On the Road to Assessing Deeper Learning: The Status of Smarter Balanced and PARCC Assessment Consortia,” CRESST Report 823 (Los Angeles: January 2013), <http://eric.ed.gov/?id=ED540597>.

⁷ James S. Coleman, “Equality of Educational Opportunity,” (Washington: U.S. Dept. of Health, Education, and Welfare, Office of Education, 1966).

⁸ J. Jennings and D. Stark Rentner, “Ten Big Effects of the No Child Left Behind Act on Public Schools,” *Phi Delta Kappan* (October 2006), http://www.pdkmembers.org/members_online/publications/Archive/pdf/k0610jen.pdf.

⁹ Linda Darling-Hammond et al., *Criteria for High Quality Assessments* (Stanford, CA: Stanford Center for Opportunity Policy in Education, 2013). See also R.C. Wei et al., “Can Performance-Based Assessments Be Reliable and Valid? Findings from a State Pilot,” presented at the annual meeting of the American Educational Research Association in Vancouver, British Columbia, April 17, 2012.

¹⁰ Richard G. Lomax, et al. “The Impact of Mandated Standardized Testing on Minority Students,” *The Journal of Negro Education* 64, no. 2 (Spring 1995): 171-85..

¹¹ Darling-Hammond et al., *Criteria for High Quality Assessments*.

¹² US Government Accountability Office, “*No Child Left Behind Act: Enhancements in the Department of Education’s review process could improve state academic assessments*, GAO-09-911 (Washington, DC: GAO, 2009), p. 20.

¹³ Organisation for Economic Cooperation and Development *OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes—Synthesis Report: What Have We Learned?* (Paris: OECD, 2013); L. Darling-Hammond and L. Wentworth, *Benchmarking Learning Systems: Student Performance Assessment in International Context*. (Stanford, CA: Stanford University, Stanford Center for Opportunity Policy in Education, 2010).

¹⁴ Ibid. See also Rosann Tung and Pam Stazesky, *Including Performance Assessments in Accountability Systems: A Review of Scale Up Efforts* (Boston: Center for Collaborative Education, 2010).

¹⁵ Measured Progress, Commonwealth Accountability and Testing System: 2007–08 Technical report, Version 1.2 (2009), retrieved February 20, 2010, from

<http://www.education.ky.gov/KDE/Administrative+Resources/Testing+and+Reporting+/Kentucky+School+Testing+System/Accountability+System/Technical+Manual+2008.htm>.

¹⁶ Jennings and Stark Rentner, “Ten Big Effects.”

¹⁷ Michael Leachman and Chris Mai, “Most States Funding Schools Less Than before the Recession” (Center on Budget and Policy Priorities, May 20, 2014), <http://www.cbpp.org/cms/?fa=view&id=4011>.

¹⁸ For examples of the consequences of a lack of coherence, see S. H. Fuhrman, *Designing Coherent Education Policy: Improving the System* (San Francisco: Jossey-Bass Publishers, 1993).

¹⁹ Linda Darling-Hammond et al., “Accountability for College and Career Readiness: Developing a New Paradigm,” *Education Policy Analysis Archives* 22, no. 86 (August 2014): 5.

²⁰ Jamal Abedi, “Performance Assessment for English Language Learners,” (Stanford Center for Opportunity Policy in Education, 2010), <https://edpolicy.stanford.edu/publications/pubs/114>.