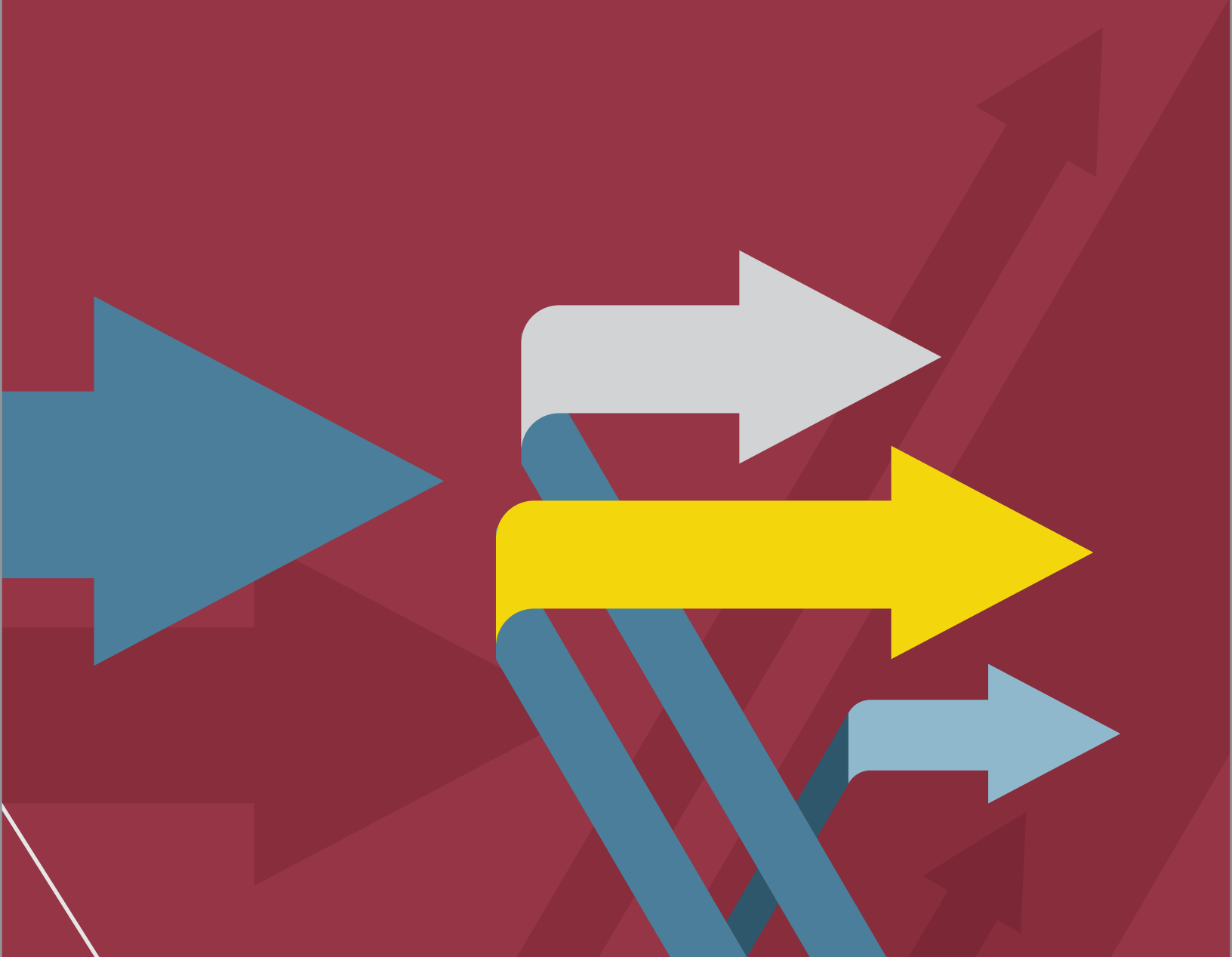


JULY 2019  
UPDATED VERSION

# The State of Assessment

A Look Forward on Innovation in State Testing Systems



Bonnie O’Keefe and Brandon Lewis



## Table of Contents

*Click on each title below to jump directly to the corresponding section.*

Introduction	4
Interim Assessments for Accountability	9
Formative Assessments to Support Instruction	13
Flexible Collaboration and Shared Item Banks	16
Science and Social Studies Exams	19
Conclusion	25
Endnotes	27
Acknowledgments	31
About the Authors	32
About Bellwether Education Partners	32

## Introduction

The past few years in state assessment have been rough. The decade began with the Obama administration's Race to the Top Assessment (RTT-A) grant, which funded states to develop higher-quality and more rigorous assessments aligned to the newly adopted Common Core Standards in math and reading.<sup>1</sup> Two multistate consortia focused on math and reading assessment kicked off their work around 2010: the Partnership for Assessment of Readiness for College and Careers (PARCC) and the Smarter Balanced Assessment Consortium (SBAC).<sup>2</sup> 45 states initially signed on. The consortia tests achieved many of their aims. Multiple studies have found that the tests align well to college- and career-ready standards, giving educators, families, and policymakers a more honest measure of student performance and progress.<sup>3</sup> The consortia also pushed the field forward on test administration and test development processes, including technological innovations that were big advancements from most tests that came before.<sup>4</sup> And comparable student test results across states were available to millions of families for the first time. Despite these achievements, by the time new tests debuted in 2015 they already faced intense backlash from political actors and later the public.<sup>5</sup> Today 12 states remain in the Smarter Balanced consortia and PARCC has essentially disbanded, although several states still administer the PARCC test or use PARCC items in their new tests.<sup>6</sup>

There were various factors behind the backlash, most of them unrelated to the quality or specific features of states' new tests. Some teachers and families pushed back on time spent testing and the perceived high stakes tied to tests, especially in states intending to

use test scores as a component in teacher evaluations.<sup>7</sup> Computer-based tests spurred investments in school technology, but they also introduced new administrative hurdles and costs. Although annual state tests had been federally required since 2001, and the consortia and standards were led by states, the new tests became a focal point of narratives about federal overreach and over-testing.

*Current wisdom holds that testing has become politically toxic. There are real risks that some states are rolling back advancements in test quality, accessibility, and rigor in the name of reducing test time and cost, or answering political pressures.*

Current wisdom holds that testing has become politically toxic. There are real risks that some states are rolling back advancements in test quality, accessibility, and rigor in the name of reducing test time and cost, or answering political pressures.

But that is not the whole story of state assessment today. In fact, there are several areas with evidence of improvement, innovation, or interesting new developments, several of which go beyond states' federally mandated role in testing. States are continuing to rethink their roles in assessment and their assessment systems in ways that may benefit teaching, learning, transparency, and equity. One of our primary goals for this brief is to inform and reinvigorate public conversation around assessment innovation at the state level, especially among education policy leaders who care deeply about educational progress, equity, and innovation, but might not see much cause for optimism or investment in state assessment.

One encouraging high-level trend in state assessments is an increasing emphasis on instructional relevance and resources that can help bring standards to life in the classroom, often as a complement to summative tests. Whereas once the state role in assessment was almost entirely limited to developing and administering traditional summative tests, states are thinking about ways to build more comprehensive assessment systems that include different kinds of tests and align with parallel efforts to improve instruction, professional development, standards, and curriculum.

New ideas in assessment may pick up steam with the help of \$17.6 million\* in competitive federal grants for assessment innovation, announced in late January 2019.<sup>8</sup> The priorities for this program include interim assessments; science, technology, engineering and math (STEM) assessments; and tests that incorporate new kinds of technical designs or project-based learning.

\*Correction: An earlier version of this document reported that the Department of Education's competitive grant for state assessments was worth \$379 million. The correct value of the grant is \$17.6 million. Congress appropriated \$378 million for state assessments in both FY 2018 and FY 2019. The majority of those funds were distributed to state agencies via formula grants. \$17.6 million was designated for competitive grants for innovation in state assessments. We have corrected this version of the report.

This brief highlights developing trends and opportunities for state systems of assessment, especially in areas beyond federally mandated reading and math assessments. Which states are pursuing these ideas, and what might be holding others back? What are the risks and rewards of investment and innovation in new test designs or assessments?

In particular we discuss the landscape of:

- 1 Interim assessments for accountability**
- 2 Formative assessments to support instruction**
- 3 Interstate collaborations and shared item banks**
- 4 Science and social studies assessments**

These four topics do not represent the totality of progress and innovation in assessment. We chose them because they are each examples where innovation and improvement is within close reach for states, and where some states are already leading in ways that may be instructive. This brief is not a comprehensive technical review, and for assessment professionals steeped in test design and administration, some topics may be familiar.

All of these ideas come with potential risks, but all are worthy of deeper investigation, discussion, and research, towards the goal of high-quality assessment systems that support equity and transparency while at the same time advancing teaching and learning. Our hope is that as a few states push beyond the status quo and show how new assessment ideas can work in the real world, more states will follow suit and create better assessment systems that support improved outcomes for students.

## Glossary

**Accountability** When we refer to accountability in this brief, we mean the system of measures, interventions, and supports for schools created by every state to meet the mandates of the Every Student Succeeds Act.

**Computer Adaptive Assessment** A computer adaptive assessment, or adaptive test, uses an algorithm to select each item for individual students dynamically based on the answers to prior questions. This kind of test design can allow for greater precision in measuring students' skills in a shorter time, because it hones in on areas where students are stronger or weaker.

**Formative Assessment** A planned, ongoing process used during teaching and learning to elicit and use evidence of student learning to improve student understanding.<sup>9</sup> These small-scale assessments (e.g., homework, exit tickets, curriculum-embedded activities) have an explicit purpose to diagnose and advance student knowledge and skills during learning.<sup>10</sup> Formative assessments are designed to play an active role in instruction, and are typically not appropriate for aggregation or use for accountability.

**Interim Assessment** A method of evaluation to measure students' knowledge and skills within a limited time frame (less than a school year). Used to inform decision-making at the classroom level and beyond. Interim assessments may be administered at set intervals, or directly after a student has learned or demonstrated knowledge in a subject area (competency-based assessments).

In this brief we explore states using multiple interims throughout an academic year for accountability purposes, replacing summative assessments.<sup>11</sup> Interim assessments can also be used as part of a formative assessment process, described below.

**Performance assessment, or performance-based tasks** Assessments that ask students to perform a task or generate a response that engages and authentically demonstrates their knowledge and skills on a topic.<sup>12</sup> Performance assessments can be short (a writing exercise) or long (a semester-long project). These kinds of assessments can be used for formative or summative purposes (or both).

**Reliability** A reliable test produces results that are consistent across students and time. If two students with the exact same level of knowledge took a reliable test, they should get the same score. A student with less knowledge should get a lower score, and a student with more knowledge should get a higher score.

**Shared Item Bank** A database of curated assessment items maintained by a collaborative (of states, districts, organizations) that can be shared and administered by multiple parties.

**Summative Assessment** Any method of evaluation that measures students' cumulative learning against standard criteria over a set period, typically over the span of the course or school year.<sup>13</sup> State summative test results are used in a variety of ways in different states and districts — as a comparable measure of student performance across schools, districts, and years; as part of state and local accountability systems to measure school performance; as a component of teacher evaluations, and in some cases as a graduation requirement or benchmark measure for students.

**Test Item** Another term for a question or task on a test.

**Validity** In educational assessment, validity refers to the accuracy of assessment results or a specific test item — whether a test accurately measures what it purports to measure, and whether the test results can support a certain use or interpretation (for example, whether a score on a state test accurately reflects student knowledge of a topic relative to their grade level, and can be used to make instructional choices for the student). The standards of validity vary depending on the intended uses of test results — a test that is used to determine whether a student graduates should have a different standard of validity than a test used to determine which students will work together on a class project.

## Innovative Assessment Demonstration Authority

One heavily discussed topic in assessment innovation is the Innovative Assessment Demonstration Authority pilot program (IADA) created under the Every Student Succeeds Act (ESSA).<sup>i</sup> This program allows up to seven states to pilot innovative state assessment models that vary from some (but not all) federal requirements for testing. States that want to participate in the pilot must ensure that their proposed assessments result in an annual summative score or proficiency determination for each student; are valid and reliable for all students, including English learners and students in special education; are comparable across districts; include a representative sample of students from around the state; and may be scaled statewide within three years. As of May 2019, the Department of Education has approved proposals from Louisiana and New Hampshire. Proposals from Georgia and North Carolina are under consideration.

When IADA was created as part of ESSA, there was considerable speculation on what states might do with increased flexibility from federal assessment requirements, especially since some state chiefs had publicly criticized ESSA's testing mandates.<sup>ii</sup> Some education analysts and assessment experts have expressed disappointment that more states did not pursue this opportunity.<sup>iii</sup> Others point to the technical constraints and requirements of the pilot, as well as the lack of additional federal funds associated directly with the waivers.

Among the four states that have applied so far, there are not many clear trends or commonalities. Because several of the ideas from IADA applications intersect with other trends of interest in this brief, such as interim assessments for accountability or social studies, Louisiana, North Carolina, and New Hampshire are discussed in greater detail below. But states do not have to participate in IADA pilots and submit to increased federal scrutiny in order to pursue innovation and improvement. The pilot program is a potential vehicle for innovative ideas, not an innovation on its own. Thus, a slow takeoff to the IADA program is not necessarily indicative of the outlook for assessment innovation.

i Office of Elementary and Secondary Education, "Innovative Assessment Demonstration Authority," US Department of Education, <https://www2.ed.gov/admins/lead/account/iada/index.html>.

ii Alyson Klein, "How Will ESSA's Innovative Assessment Pilot Work?," Education Week, June 30, 2016, [http://blogs.edweek.org/edweek/campaign-k-12/2016/06/how\\_will\\_essas\\_innovative\\_asse.html](http://blogs.edweek.org/edweek/campaign-k-12/2016/06/how_will_essas_innovative_asse.html).

iii Alyson Klein, "What Happens When States Un-Standardize Tests?," Education Week, October 11, 2018, <https://www.edweek.org/ew/articles/2018/10/11/what-happens-when-state-un-standardize-tests.html>.



## Interim Assessments for Accountability

Right now, most students take their state exams once per year, and results might not come back to schools and families until after the start of the next school year.<sup>14</sup> If one of the goals for state summative tests is to inform instruction and school decision-making, this means students do not get timely feedback on their performance, and by the time school leaders and teachers see last year's scores, instruction for the next year is already underway. Practically, this means the cycle by which state assessments can inform instruction and school decisions is slow and limited.

A chance to alter this model is already embedded in federal education law, but few states seem interested. ESSA invites states to replace their end-of-year summative assessments with assessments administered over the course of the year, also called interim assessments. States' assessments may "be administered through multiple statewide interim assessments" to provide "valid, reliable, and transparent information on student achievement or growth."<sup>15</sup>

Additionally, two of the priority areas for competitive federal assessment grants announced early in 2019 include "the development of comprehensive academic assessment instruments ... that emphasize the mastery of standards and aligned competencies in a competency-based education model," and "measuring student academic achievement using multiple measures from multiple sources."<sup>16</sup> Various models of interim assessments for accountability could meet one or both of these priorities.

*Interim assessments for accountability would be mandatory, and would replace state tests in an accountability system.*

Many schools and districts already use interim assessments for their own benchmarking and instructional planning purposes, either with state-provided tests, tests purchased from vendors, or tests designed on their own. In contrast, interim assessments for accountability would be mandatory, and would replace state tests in an accountability system.

### **Benefits**

Interim assessments for accountability offer some strong potential benefits for states, teachers, and schools. Teachers could receive actionable feedback and adjust instruction over the course of the year. School districts could save money if they no longer need to purchase supplementary interim assessments, and states could be assured of the quality of tests schools use during the year. Interim assessments are usually shorter, and because of this they may be easier to integrate into schools' day-to-day schedules. This could discourage weeks of pretest preparation and a long period of "testing mode" near the end of the year.

Some states are also seeking to influence instructional practices. If each interim assessment covers fewer standards at a deeper level, it could encourage more depth of instruction in the classroom. More flexible testing models could potentially encourage the adoption of personalized or competency-based instruction, because it could allow students and/or schools to demonstrate proficiency in a particular subset of standards on different timelines.

### **Risks**

The technical, political, and logistical challenges of designing and implementing this kind of assessment are considerable. First, influencing curriculum, scope, and sequence might be part of states' goals for this work, but it may also generate opposition. States have historically been agnostic about when during the academic year students should be able to master standards, but if states design interim assessments to test subsets of standards at particular intervals, it could reshape the pacing and ordering of classroom curriculum to align with the tests. This could be part of a state's goals for interim assessments, but it may also provoke pushback in the name of local control.

A more flexible, competency-based interim test design would have less of an impact on curriculum pacing, but could pose different risks to comparability of test data for accountability. If two schools administer components of the assessment at different points of the year, after different amounts of instructional time, in different orders, can students' scores be fairly compared?<sup>17</sup> When there are high stakes linked to an assessment, small degrees of variation or potential for bias can have a big effect on reliability and validity.

Logistically, interim assessments for accountability may actually take up more combined time and resources than traditional summative assessments. States will need to increase their capacity to provide nearly year-round assessment administration support and schools

will need to adapt to a new mode of testing in their teaching and in their administrative procedures.<sup>18</sup> Some districts may choose to stick with their own interim assessments in addition to the state test, creating redundancy.

In order to see success with this model, states must be willing to put some kind of stake in the ground around curriculum and sequencing, and work with districts and teachers to ensure that results are both useful and reliable. Interim assessments have the potential to provide data to teachers on a faster timeline and reduce costs for districts, but state policymakers should not forget that teachers and administrators will need additional supports and guidance to use these new kinds of tests and the results they generate.

*In order to see success with this model, states must be willing to put some kind of stake in the ground around curriculum and sequencing, and work with districts and teachers to ensure that results are both useful and reliable.*

## State Examples

### Nebraska

Nebraska is currently planning to develop interim assessments for accountability purposes in grades 3–8.<sup>19</sup> Nebraska aims to give students the opportunity to interact with their own data and monitor their growth throughout the year, and to provide teachers with real-time data to improve classroom instruction.<sup>20</sup> This system will include an adaptive end-of-year test and interim assessments in grades 3-8, while high school students will continue to take the ACT test.<sup>21</sup> Starting in 2017–18, Nebraska partnered with NWEA to develop the new test and offer MAP Growth optionally to all schools. MAP Growth is an adaptive interim assessment used by schools in many states across the country, which measures student skills across grade levels.<sup>22</sup> Although it is optional, all Nebraska school districts now use MAP Growth. As a next step, the state will develop a new state adaptive interim assessment system, which will measure student performance against grade-level expectations for the purposes of accountability requirements, and replace the current end-of-year test.<sup>23</sup>

### Louisiana

Louisiana was the first state to win approval under the Department of Education's IADA pilot for a new humanities assessment, which will measure students' understanding of English language arts (ELA) and social studies content with interim assessments.<sup>24</sup> This new test will align with content in Louisiana's ELA Guidebooks curriculum, an optional resource aligned with college- and career-ready standards developed by the state.<sup>25</sup> Several brief assessments including both ELA and social studies content will be administered throughout the year, rather than in end-of-year ELA and social studies exams. Tests will focus on books or passages that students have studied in class.<sup>26</sup> As a result, state leaders hope that teachers can focus instruction on important background knowledge and making meaning of full texts, and students will have equitable opportunities to develop knowledge about topics covered on the assessments and in the standards.<sup>27</sup>

## *North Carolina*

North Carolina recently submitted an IADA pilot application to expand an existing interim assessment approach, called NC Check-Ins, into a system that could generate an end-of-year score and be used for accountability.<sup>28</sup> These assessments would be administered three to four times per year in flexible test windows, and use items directly drawn from North Carolina's existing end-of-year tests. Under the proposal, the state would still maintain an end-of-year test for students who missed earlier interim tests, and potentially as a retest option for students who do not demonstrate proficiency in the interim assessments. Reducing test time and increasing transparency for students, teachers, and families are among the state's top goals for this proposal. North Carolina's application is still under consideration as of April 2019, and federal reviewers requested more information from the state on a variety of topics, including alignment between the interim assessment and state standards, how the state will arrive at a summative score based on interim tests, and plans for eventually scaling the pilot statewide.<sup>29</sup>

### Sidebar 2

## Innovation in Assessments for Special Populations

Federal grant programs in the past decade not only funded PARCC and Smarter Balanced — they supported four additional multistate consortia focused on higher-quality assessments for specific student groups. DLM and MSAA/NCSC created new alternate assessments for students with significant cognitive disabilities. WIDA ACCESS and ELPA21 created new assessments of English language proficiency. Many states still administer all of the above tests, and they have not experienced the politicization and backlash around the higher-profile math and reading assessments.

This brief does not focus on these tests in detail, but one of the priorities of the new competitive federal assessment innovation grants is innovation in tests for both special education students and English learners. With these additional resources, it is possible that assessments for special populations of students could be a new frontier in assessment innovation and advancement.

## Formative Assessments to Support Instruction

An increasing number of states are using guidance, training, and incentives to shape formative assessments for instructional purposes. Formative assessments are a planned, ongoing process of low-stakes, informal assessment used by students and teachers during learning.<sup>30</sup> Formative assessments are not separate from instruction, but are opportunities for students to practice what they have learned during commonplace classroom activities and opportunities for teachers to adjust their instruction based on quick analysis of student performance.<sup>31</sup>

*One of the things state leaders heard loudly from the pushback to testing was that state summative assessments were rarely useful to teachers directly.*

These kinds of lower-stakes assessments have not traditionally been a part of states' assessment purview, as they do not generate data that is appropriate for use in accountability or for aggregation. But one of the things state leaders heard loudly from the pushback to testing was that state summative assessments were rarely useful to teachers directly.<sup>32</sup> Other evidence suggested that many schools lacked high-quality instructional materials and practices that could help students build the skills and knowledge they needed to succeed on more rigorous state tests.<sup>33</sup>

In response, more states are getting involved in formative assessments for instruction by providing or recommending vetted assessment resources aligned to state standards. States are offering guidance, purchasing resources, or subsidizing the cost of a menu of

assessment options.<sup>34</sup> Some states' increased activity around formative assessments for instruction kicked off with the optional interim assessments made available by the Smarter Balanced consortium.<sup>35</sup> Positive feedback around those resources may have generated interest in more state-provided low-stakes or no-stakes tests and tools.<sup>36</sup> Additional resources around formative assessment for instruction can be especially useful in pre-K through second grade or in subjects where summative tests are not required by law.

### Benefits

*The demand and interest in higher-quality formative assessments for instruction among schools and districts is clear in the growing market for classroom assessments.*

The demand and interest in higher-quality formative assessments for instruction among schools and districts is clear in the growing market for classroom assessments. In 2017, spending on classroom assessments reached nearly \$1.6 billion nationally, compared to \$1.3 billion on state-level mandatory tests.<sup>37</sup> States may have more capacity than districts to differentiate among options in the market or design custom resources, evaluate quality and alignment with state standards, and subsidize higher-quality choices.

If states offer these kinds of resources in addition to high-quality tests that can be used for accountability, it could help more teachers and school leaders improve instruction and understand college- and career-ready standards, ultimately helping more students reach those standards. It could also help build trust among state agencies, districts, and educators if the state provides resources around assessment that are exclusively meant for instructional, not accountability purposes.

### Risks

States will need to strengthen their internal capacity to effectively do work that was historically outside of their purview. To address this concern, states should consider partnerships with local universities or professional associations that could enhance the design and implementation of formative assessment resources. States will also need to invest in training and professional development to help administrators and teachers put the new resources at their disposal to use in ways that will serve students.

There is also a risk that educators will associate any test that comes from the state with high stakes and accountability, and thus mistrust it. This is especially salient if student data are collected in a computer-based format. These assessments will not achieve their goals if teachers and school leaders are resistant to using them. States can mitigate this risk through clear communication, outreach, and training about the intended purpose and uses of various assessments.

*Given tight budgets in some states, it is critical that investments in formative tests for instruction do not supplant efforts to maintain and improve other tests that can be used for public reporting, equity monitoring, and accountability systems.*

Given tight budgets in some states, it is critical that investments in formative tests for instruction do not supplant efforts to maintain and improve other tests that can be used for public reporting, equity monitoring, and accountability systems. These kinds of tests serve different, complementary purposes, and one should not come at the expense of the other.

### **State Examples**

#### *Center on Standards & Assessment Implementation Consortium*

The Formative Assessment Bi-Regional Advisory Board (FAB-RAB) is a consortium of seven states committed to incorporating formative assessment as part of their comprehensive assessment system.<sup>38</sup> Through partnership with the WestEd/CRESST Center on Standards and Assessment Implementation, sponsored and facilitated by the Center and South Central Comprehensive Centers at the University of Oklahoma, participating states receive professional development for state leaders, discuss and resolve common problems associated with formative assessment implementation at the district and classroom level, create common regional resource materials, and collaboratively develop long-range implementation plans.<sup>39</sup> Member states have initiated their own formative assessment advisory boards, and have focused on ways to embed formative assessment practices into instruction across their states.

#### *Michigan*

Formative Assessment for Michigan Educators (FAME)<sup>40</sup> is a joint effort of the Michigan Department of Education and the Michigan Assessment Consortium that trains Michigan teachers, coaches, and leaders in effective formative assessment processes.<sup>41</sup> FAME not only offers trainings and resources, it organizes teachers and coaches into a system of “Learning Teams” that sustain the work of formative assessment in schools. Educators apply to the state, build a local “Learning Team,” and receive initial and ongoing training on FAME’s model of formative assessment and peer support. Regional staff support the educator teams on an ongoing basis. Several research studies of the FAME model are underway.<sup>42</sup>

## Flexible Collaboration and Shared Item Banks

*Item banks are not a new development in assessment. But states are participating in these banks and collaborating on tests in new ways, especially as the number of states participating in consortia tests has dwindled.*

In the PARCC and Smarter Balanced models, states co-developed and shared whole assessments. But every test is composed of questions (also called “items”), each of which can cost hundreds or thousands of dollars to develop and validate. Instead of sharing whole tests, more states now mix, match, and share large groups of items via shared item banks. An item bank is a database of test items, which can be owned by a vendor, a nonprofit organization, a state, or a group of states. Item banks are not a new development in assessment. But states are participating in these banks and collaborating on tests in new ways, especially as the number of states participating in consortia tests has dwindled. Shared item banks are now common practice in large-scale assessment design, not only in math and reading, but in science, social studies, alternative assessments, and English language assessments.

Federal law required that any items developed with federal dollars, including items developed under the RTT-A grants, be made “freely available to states, technology platform providers, and others that request it for purposes of administering assessments.”<sup>43</sup> This does not mean the items created by the consortia are free, but RTT-A grants created a plethora of new, high-quality items aligned to college- and career-ready standards that were available to consortia and non-consortia members alike. As states sought out a way to create new tests of their own, using items from multiple shared sources in addition to state-created models was an appealing option.



## Benefits

There are many benefits to cross-state collaboration and shared item banks, which is part of the reason why they have become common practice in state assessments. The top reasons to do so are high-quality items at lower cost, and more cross-state brainpower and idea sharing. Technical details such as the best practices for agreements among states and vendors involved in item banks are fairly well-established in the field.

There are costs associated with purchasing items from an item bank, but these are generally much lower than making new tests from scratch because multiple users of an item bank can share costs. In some cases, states pay back to the item bank in kind by contributing their own items to be vetted and used by others.

*Beyond cost-sharing benefits, this model is cause for optimism because it can support higher-quality tests and interstate collaboration even as states shift away from the consortia.*

Beyond cost-sharing benefits, this model is cause for optimism because it can support higher-quality tests and interstate collaboration even as states shift away from the consortia. Shared item banks and hybrid test designs may also give states the ability to maintain stability and comparability in their scores over time, even if their tests change.<sup>44</sup> For instance, if a former PARCC state uses enough PARCC items in its new test, it could establish links between the two tests that allow it to track growth trends over time.

Theoretically, a similar approach could also be used for some kinds of cross-state comparisons, even if states are not administering the exact same tests. Over time, it is possible that these collaborations around items could evolve into more ambitious cross-state endeavors that would support new kinds of innovative thinking around test design.<sup>45</sup>

But states do not have to be aiming for comparability in order to participate in item bank collaborations. One advantage of this model is that states can maintain as much autonomy over test design decisions as they desire.

## Risks

There are relatively few technical risks to participating in a shared item bank, as long as states set clear, high standards for item quality, and consider issues such as equitable cost sharing among states, test item security, and agreements on timelines for releasing items to the public. As explained above, because various kinds of shared item banks have quickly become common, most states and vendors have developed reliable strategies for addressing these questions in cross-state agreements and planning stages. These kinds of agreements were put to the test on a wide scale with PARCC and Smarter Balanced.

States face bigger risks going it alone. Tennessee is one such cautionary tale. After the state legislature abruptly pulled out of the PARCC consortium, the state tried to quickly develop its own assessment system, TNReady. The state spent more than \$130 million in hasty test development only to be plagued by years of technical problems and vendor failures in ways that negatively impacted the usability of their results.<sup>46</sup>

The less predictable risks to cross-state collaboration are primarily political, especially if states are aiming for cross-state comparability in their results. If political winds shift, support for collaboration across state lines may shift as well.

### **State Examples**

#### *New Meridian*

In 2017, New Meridian, an assessment design and development organization, was selected as the authorized agent of the Council of Chief State School Officers (CCSSO) to maintain the item bank and test designs developed by the former PARCC states. States can license some or all of that content from New Meridian for their own tests, along with options for various kinds of advisory and test development support.<sup>47</sup> New Meridian's approach developed in direct response to states' need for high-quality assessment content at a reasonable cost with autonomy around design decisions, especially among states leaving the consortia.<sup>48</sup> So far many of New Meridian's partner states are former PARCC states.<sup>49</sup> States still have access to some of the advantages of a consortium (high-quality items at a reduced cost), and continuity in their test design and content, with the flexibility to make individual choices. New Meridian's current focus is producing shorter and/or adaptive assessment designs for states that build upon the content and design of PARCC.<sup>50</sup>

#### *Michigan and Smarter Balanced*

In 2014, the Michigan legislature required the Michigan Department of Education (MDE) to develop a new state test for spring 2015. As a result Michigan rebranded Smarter Balanced ELA and math content while simultaneously developing new content aligned to state standards in one assessment called the Michigan Student Test of Educational Progress (M-STEP).<sup>51</sup> The current assessment is a hybrid of performance tasks, classroom activity, and an online assessment that includes both Smarter Balanced and Michigan-developed content.<sup>52</sup>

## Science and Social Studies Exams

*Science and social studies content are both especially well-suited to performance-based, hands-on activities and tasks, which can be engaging for students and provide deeper information on learning to teachers.*

Science and social studies assessments are ripe for state innovation and experimentation. Generally, state action in these content areas requires a committed state leader or other driving force because they are mostly optional, costly, and usually do not count towards things like accountability ratings. But these subjects could be a laboratory of assessment innovation, with wide flexibility in law and subject areas that have been somewhat neglected in the era of math and reading assessments. Science and social studies content are both especially well-suited to performance-based, hands-on activities and tasks, which can be engaging for students and provide deeper information on learning to teachers.<sup>53</sup>

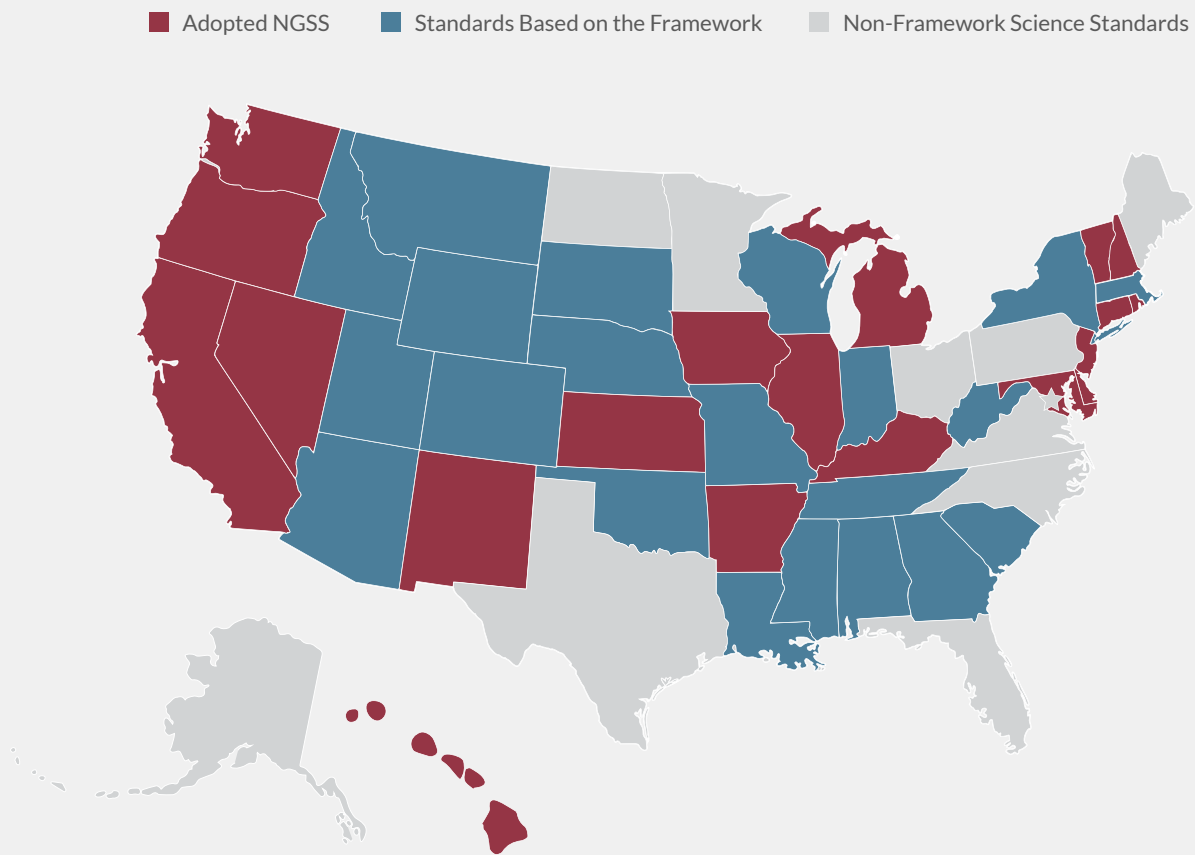
Science and social studies assessments have different policy contexts from math, reading, and each other. ESSA requires science exams once in each grade band.<sup>54</sup> ESSA does not require assessments in social studies, but 33 states offer some kind of social studies assessment. Neither science nor social studies assessments are required components of state accountability systems, although some states choose to use them.

Some states have cut back non-federally-required exams in these subjects in response to over-testing pressure. But advocacy groups, educators, and experts in these subject areas often want some kind of comparable assessment because it affirms the importance of the subject area, provides valuable performance and progress data, and helps protect against cutbacks in time or resources in favor of tested subjects.<sup>55</sup> This also has equity implications: Students in low-income schools and students of color are disproportionately likely to experience cuts to science and social studies resources.<sup>56</sup>

## Science

States have not traditionally invested much time and resources into their science tests compared with math and reading, a challenge that is compounded by many states' adoption of new, more rigorous and complex science standards in recent years.<sup>57</sup> 40 states have adopted new science standards in the past five years, inspired in large part by the influential "Framework for K–12 Science Education," released in 2012 by the National Research Council.<sup>58</sup> Following the release of the Framework, the National Academies of Science, science educators, and Achieve collaborated to develop the Next Generation Science Standards (NGSS).<sup>59</sup> 19 states have adopted the NGSS, while 21 states have adopted standards inspired by the Framework.<sup>60</sup>

Figure 1 K–12 Science Standards Adoption



Source of Standards Adoption Information: NGSS, <https://ngss.nsta.org/About.aspx>.

The NGSS, and other Framework-based science standards, differ in many ways from prior state standards. NGSS emphasizes learning and applying science concepts in the context of real-world scientific phenomena and events.<sup>61</sup> The defining feature of the NGSS is its three interlocking core components: Disciplinary Core Ideas,<sup>62</sup> Cross Cutting Concepts,<sup>63</sup> and Science & Engineering Practices.<sup>64</sup> Each NGSS standard includes all three core components, creating a “three-dimensional standard.”

For example, a third-grade standard in the NGSS asks that students “obtain and combine information to describe climates in the world.”<sup>65</sup> This single standard includes a science and engineering component on obtaining and evaluating information from various sources, core ideas about understanding weather and climate, and a cross-cutting concept that patterns in nature can be observed and used to make predictions.

It is possible to reliably assess a standard like this in a test format that stays true to the intent of the standards. After all, math and ELA standards are also complex, but methods for test and item design aligned with rigorous new standards are now fairly well established. But science tests have not had comparable levels of investment and time from states, and have not played the same role in accountability systems.

*There are signs that some states are seeking to create better science tests, and investing in instructional supports aligned to the rigor of new standards.*

There are signs that some states are seeking to create better science tests, and investing in instructional supports aligned to the rigor of new standards. State efforts in science include cross-state collaboration on test items and design, hands-on tasks that can be used in class, and instructional resources that lie outside the summative test. Science tests could get a new infusion of support from the U.S. Department of Education: 2019 grants will favor proposals with a Science, Technology, Engineering and Math (STEM) focus, especially in computer science.<sup>66</sup> Thus far, computer science assessments have mostly been the purview of the AP, but the incentives of available federal funding could galvanize states that have already demonstrated interest and investment in this subject area.

### **Benefits and Risks**

The primary tradeoffs in science assessment revolve around cost, technology, and comparability. Technology could enable more ambitious, interactive assessments that use computerized simulations to create performance-based science tasks.<sup>67</sup> The more realistic and rich simulations are also expensive to develop and complex to score. Tests that meld the full interactive capabilities of modern video games, the depth of the science standards, and the reliability of a traditional standardized test are likely still several years in the future; however, the components to build them exist today.<sup>68</sup>

Live, performance-based tasks in the classroom are also an attractive option for science assessments, but these would still require research and careful training for teachers in order to yield data that could be useful, valid, and reliable for accountability and instruction.

## State Examples

### *New Hampshire*

New Hampshire has been a leading innovator in comparable, performance-based assessments through its Performance Assessment of Competency Education (PACE) system, which spans math, reading, and science. PACE began more than five years ago, but was recently approved under ESSA's IADA pilot. PACE integrates locally designed performance assessments with more traditional tests in some grade levels. In science, students take performance assessments in grades three, four, six, seven, and eight. In fifth grade, students take a summative science test. In high school they take both a summative test and course-specific performance assessments. New Hampshire's model demonstrates both the opportunities and the challenges of integrating live performance-based tasks into an assessment system for accountability purposes. The state has not scaled PACE statewide, but it has developed a tiered system by which districts can gradually transition into PACE, and plans to expand further in future years.

### *Kentucky*

Kentucky was an early adopter of NGSS and one of the first states to publicly pursue new science assessment development. Kentucky's approach combines elements of interim test design with performance-based tasks. As an example, students in high school are asked to analyze the data from trials of a test dummy colliding with various airbag designs, and then prepare a written argument for which airbag provides the most protection.<sup>69</sup> The intent of this design is to allow students to demonstrate standards skills and knowledge in class, as they learn the material, often at a deeper level than is possible on a traditional standardized assessment.<sup>70</sup>

### *OpenSciEd*

OpenSciEd is a collaborative effort among 10 states, curriculum developers, and science experts to create an open-source curriculum aligned to the NGSS and the Framework, answering the demand for high-quality instructional materials.<sup>71</sup> The ultimate results will likely include integrated formative assessments and other kinds of instructional resources that might inform the shape of better summative science tests.

## Social Studies

Unlike science, social studies is not required by ESSA, and has not gone through the same transformative multistate standards experience. Despite the significant number of states with a social studies test requirement of some sort, innovation and investment seem to be largely absent, which is a major missed opportunity for those who want to lift up the value of social studies instruction, and encourage progress and equity in social studies.

Table 1 > Social Studies Assessments in States<sup>72</sup>

U.S. Citizenship Test (required to receive HS diploma)	K-8 Assessment	U.S. History & Government
8 States AL, AZ, AK, ID, MN, ND, TN, UT	13 States CO, DE, GA, IN, KS, KY, LA, MI, SC, TN, TX, VA, WI	12 States FL, GA, KY, LA, MI, MO, NY, OH, OK, SC, TN, TX

Recent trends in social studies assessment have mostly been to cut state-developed tests, and in eight states, to introduce the U.S. Citizenship Test as a high school requirement. The citizenship test trend was backed in part by a major push by the Civics Education Initiative, whose leaders were concerned by the de-emphasis of civics education in schools and Americans' low level of basic knowledge about government and civic engagement.<sup>73</sup> Social studies teachers and advocacy groups have pushed back against this trend, claiming that the test is not a high-quality representation of the skills and knowledge needed to develop a comprehensive view of civics. The citizenship test only measures students' ability to memorize discrete facts, and does not require meaningful engagement with the depth or rigor of social studies content standards. Others have argued that the test contains factual errors.<sup>74</sup>

*Without accountability pressures, social studies could be a testing ground for new kinds of performance-based tasks, growth models, or class-embedded assessments.*

### Benefits and Risks

Without accountability pressures, social studies could be a testing ground for new kinds of performance-based tasks, growth models, or class-embedded assessments. However, these kinds of tests may not be able to serve the same uses for tracking performance publicly and understanding achievement gaps as tests that are subject to more stringent requirements. A high-quality in-class social studies assessment might support better learning outcomes for teachers who choose to use it, but there is no way to say for sure whether students are progressing without any public reporting.

## State Examples

### *Louisiana and Washington*

There are a handful of states doing interesting things with their social studies tests. As described earlier in this brief, Louisiana is piloting an innovative assessment that integrates social studies content with ELA skills. Washington is another notable state in social studies test development, but its tests exist in a gray area: Students must take social studies assessments in several grades, but districts may choose to use the state tests or develop their own, which do not have to be comparable. The state's test is a multistep project-based social studies and civics assessment aligned to state standards, designed to push students' critical thinking skills.<sup>75</sup> These assessments also have connections to Washington's English Language Arts standards so that teachers are encouraged to integrate reading and writing tasks in social studies instruction, and vice versa.<sup>76</sup> The state legislature recently passed a citizenship test requirement in addition to existing assessments, which will take effect in 2020 as part of a broader emphasis on civic inquiry, engagement, and action.

## Science and Social Studies Assessment in the Future

Both science and social studies assessments are ripe for innovation and improvement. If policymakers are frustrated that their assessment ideas — be it performance-based assessments, curriculum-embedded tests, or locally designed assessment — are not immediately feasible for math and reading, the field of possibility for science and social studies is wide open. Success designing new kinds of tests in these topic areas could also provide an evidence base to support similar changes in math and reading, and encourage new kinds of deep thinking about the relationship between high-quality assessment and high-quality instruction.



## Conclusion

**A**cross the topic areas above, states tended to fall into one of three camps. A small number of trailblazing states are making big, public plays around innovation in assessment. This group includes states that are out front in the innovative assessment demonstration pilots, and those aggressively pursuing investments in areas such as interim assessments for accountability. These states are committing significant resources to new models and ideas, and pushing the boundaries of what a state assessment system could look like.

The second group of states are shifting their role in assessment more gradually, and considering new ways to simultaneously advance equity, transparency, and teaching and learning. This group is also exploring new ways for states to engage in assessment and curriculum outside of their traditional lanes, although change has been slower and less consistent across topic areas. The trailblazers are few in number, but the second group, the shifters, may be larger than we know, suggesting a somewhat brighter future for state assessments than public discourse might suggest.

A third group of states are in retreat on assessment, reducing any investment to the barest minimum. Some of these states may be at risk of compromising test quality in the name of cutting test time or responding to political pressures. This could leave students worse off, with less accurate information and transparency around equity gaps and long-term college- and career-readiness goals, and reverse the progress made over the past decade. More ambitious states may serve as models and proof points to reverse this trend.

*Some of these states may be at risk of compromising test quality in the name of cutting test time or responding to political pressures.*

Greater attention to a positive role for assessment in instruction alongside accountability and equity purposes for testing could help avert the backlash seen in earlier efforts. More importantly, this renewed focus on state supports for standards-aligned instruction could help fulfill the promise of more rigorous college- and career-ready standards: that if states set a higher bar, more students will reach it. State movements towards innovative assessments parallel growing state involvement in encouraging high-quality curricula, facilitating more effective forms of instructional coaching and professional learning, and other encouraging efforts to translate states' goals for students into classroom-level supports and change.

The specific topic areas explained above intersect with one another in unexpected ways. We see examples of science assessments created by informal consortia, interim assessments for accountability that fold in formative and instructional resources, and innovative assessment proposals that bring interdisciplinary classroom instruction to the forefront. Over \$17 million in competitive grants that the federal government plans to award in summer 2019, with priority areas that intersect with many of these ideas, could also reshape the field.

Of course, not everything states could do is something they should do. When millions of dollars and precious student learning time are at stake, states must weigh the potential risks of change. Stability may be especially valuable among states that have experienced three or more assessments in as many years. New innovations and ideas, especially in areas of assessment that are not in states' typical responsibilities, should not jeopardize previous progress on accountability, equity, and transparency.

This work can be extremely difficult. States embarking on this path should have clear goals, a theory of action centered on equity and student learning, and realistic expectations for the substantial resources, capacity, and time it takes to research, develop, implement, and sustain a comprehensive system of high-quality assessments. They also should not underestimate the importance of proactive, clear communications with districts, educators, families, and political leaders outside of education.

The opportunities for assessment improvement available under ESSA, with or without any federal permission slip, should be something that states seriously consider. States are building their goals, strategies, and accountability systems on assessments, and using them to guide instructional decisions or factor in to graduation and student promotion. Seeking continuous improvement and innovation in this area is important and urgent because of the ongoing impact and importance of high-quality assessment systems for students, teachers, and our educational system as a whole.

*The opportunities for assessment improvement available under ESSA, with or without any federal permission slip, should be something that states seriously consider.*

## Endnotes

- 1 Office of State Support, "Race to the Top Assessment Program," US Department of Education, <https://www.ed.gov/open/plan/race-top-assessment>.
- 2 Office of State Support, "Race to the Top Assessment Program Awards," US Department of Education, <https://www2.ed.gov/programs/racetothetop-assessment/awards.html>.
- 3 Sheila Schultz et al., "Evaluating the Content and Quality of Next Generation High School Assessments," HumRRO, 2016, [https://www.humrro.org/corpsite/sites/default/files/HQAP\\_HumRRO\\_High\\_School\\_Study\\_Final%20Report.pdf](https://www.humrro.org/corpsite/sites/default/files/HQAP_HumRRO_High_School_Study_Final%20Report.pdf); Ira Nichols-Barrer et al., "Predictive Validity of MCAS and PARCC," 2015, Mathematica, <https://www.mathematica-mpr.com/our-publications-and-findings/publications/predictive-validity-of-mcas-and-parcc-comparing-10th-grade-mcas-tests-to-parcc-integrated-math-ii>; Nancy Doorey and Morgan Polikoff, "Evaluating the Content and Quality of Next Generation Assessments," Fordham Institute, 2016, <https://files.eric.ed.gov/fulltext/ED565742.pdf>.
- 4 Samantha Batel and Scott Sargrad, "Better Tests, Fewer Barriers: Advances in Accessibility through PARCC and Smarter Balanced," Center for American Progress, February 25, 2016, <https://www.americanprogress.org/issues/education-k-12/reports/2016/02/25/131789/better-tests-fewer-barriers/>; Scott Marion, "Bursting the Bubbles in American Testing," FutureEd, 2017, <https://www.future-ed.org/bursting-the-bubbles-in-american-testing/>.
- 5 Ashley Jochim and Patrick McGuinn, "The Politics of the Common Core Assessments: Why States Are Quitting the PARCC and Smarter Balanced Testing Consortia," Education Next, fall 2016, <https://www.educationnext.org/the-politics-of-common-core-assessments-parcc-smarter-balanced/>.
- 6 Catherine Gewertz, "Big Things You Need to Know Now About the PARCC and Smarter Balanced Tests," *Education Week*, 2019.
- 7 Amanda Paulson, "Standardized Test Backlash: More Parents Pull Kids from Exams as Protest," *The Christian Science Monitor*, April 30, 2015, <https://www.csmonitor.com/USA/Education/2015/0430/Standardized-test-backlash-More-parents-pull-kids-from-exams-as-protest>.
- 8 Office of Elementary and Secondary Education, "Applications for New Awards; Competitive Grants for State Assessments Program," US Department of Education, January 28, 2019, <https://www.federalregister.gov/documents/2019/01/28/2019-00217/applications-for-new-awards-competitive-grants-for-state-assessments-program>.
- 9 Formative Assessment for Students and Teachers (FAST) State Collaborative on Assessment and Student Standards (SCASS), "Revising the Definition of Formative Assessment," Council of Chief State School Officers, 2018, <https://ccsso.org/resource-library/revising-definition-formative-assessment>.
- 10 Perie et al., "The Role of Interim Assessments in a Comprehensive Assessment System."
- 11 Marianne Perie et al., "The Role of Interim Assessments in a Comprehensive Assessment System," Achieve Inc., The Aspen Institute, and Center for Assessment, 2007, <https://www.achieve.org/files/TheRoleofInterimAssessments.pdf>.
- 12 School Redesign Network, "What Is Performance-Based Assessment?," Stanford University, 2008, <https://edpolicy.stanford.edu/sites/default/files/events/materials/2011-06-linked-learning-performance-based-assessment.pdf>.
- 13 Abigail Swisher, "Recipe For Success: What State Lawmakers Need to Know About K-12 Accountability Assessments," New America, 2017, <http://na-production.s3.amazonaws.com/documents/Recipe-for-Success.pdf>.
- 14 John Fensterwald and Theresa Harrington, "Smarter Balanced Scores Could Be Delayed Two More Weeks," EdSource, August 29, 2017, <https://edsources.org/2017/california-delays-release-of-smarter-balanced-scores/586404>.
- 15 Every Student Succeeds Act, Public Law 114-95, §1111(b)(2)(B)(viii).
- 16 US Department of Education, "Applications for New Awards: Competitive Grants for State Assessment Program," *Federal Register* 84, no. 18, January 28, 2019, <https://www.govinfo.gov/content/pkg/FR-2019-01-28/pdf/2019-00217.pdf>.

- 17 Nathan Dadey and Brian Gong, "Using Interim Assessments in Place of Summative Assessments? Consideration of an ESSA Option," Council of Chief State School Officers, 2017, [https://ccsso.org/sites/default/files/2017-12/ASR\\_ESSA\\_Interim\\_Considerations-April.pdf](https://ccsso.org/sites/default/files/2017-12/ASR_ESSA_Interim_Considerations-April.pdf).
- 18 Ibid.
- 19 "Interim and Formative Assessment Supports," Nebraska Department of Education, <https://www.education.ne.gov/assessment/interim-and-formative-assessments/>.
- 20 Ibid.
- 21 Sean Cavanagh, "NWEA Poised to Capture Statewide Testing Contract in Nebraska," Ed Week Market Brief, May 31, 2017, <https://marketbrief.edweek.org/marketplace-k-12/nwea-poised-win-statewide-testing-contract-nebraska/>.
- 22 Computer adaptive tests use an algorithm to select harder or easier questions for students based on their answers to prior questions. This can allow the test to measure student knowledge of a topic with fewer questions. Smarter Balanced, MAP, and several other state tests use a computer adaptive model.
- 23 "Interim and Formative Assessment Supports," Nebraska Department of Education.
- 24 Louisiana Department of Education, "Louisiana Believes: Louisiana's Application for Innovative Assessment Demonstration Authority under Section 1204 of the Elementary and Secondary Education Act (ESEA)," US Department of Education, 2018, <https://www2.ed.gov/admins/lead/account/iada/laiadaapplication.pdf>.
- 25 "ELA Guidebooks 2.0," Louisiana Department of Education, <https://www.louisianabelieves.com/academics/ela-guidebooks>.
- 26 Ibid.
- 27 Future Ed, "Strengthening Standards, Teaching and Tests in a Deep Red State," Georgetown McCourt School of Public Policy, January 2, 2019, <https://www.future-ed.org/reforming-standards-curriculum-and-testing-in-a-deep-red-state/>.
- 28 North Carolina Department of Education, "North Carolina Application for New Authorities under the Innovative Assessment Demonstration Authority," US Department of Education, December 14, 2018, [https://content.govdelivery.com/attachments/NCSBE/2018/12/18/file\\_attachments/1125667/NorthCarolina\\_IADA\\_Final\\_121418\\_version%202.pdf](https://content.govdelivery.com/attachments/NCSBE/2018/12/18/file_attachments/1125667/NorthCarolina_IADA_Final_121418_version%202.pdf).
- 29 Frank Brogan, "North Carolina Interim Feedback Letter," US Department of Education, March 8, 2019, <https://www2.ed.gov/admins/lead/account/iada/nciadainterimfeedback03082019.pdf>.
- 30 FAST SCASS, "Fundamental Insights about Formative Assessment," Council of Chief State School Officers, October 2018, <https://ccsso.org/sites/default/files/2018-10/FAST%20SCASS%20Fundamental%20Insights%20about%20Formative%20Assessment.pdf>.
- 31 Ibid.
- 32 Andrew Latham, "ESSA and the New Era of Local Control for Assessment and Accountability," in "The Road Ahead for State Assessments: What the Assessment Consortia Built, Why it Matters, and Emerging Options," Educational Testing Service, 2016, [https://www.ets.org/s/k12/pdf/coming\\_together\\_the\\_road\\_ahead.pdf](https://www.ets.org/s/k12/pdf/coming_together_the_road_ahead.pdf)
- 33 Michele Molnar, "Market Is Booming for Digital Formative Assessments," *Education Week*, May 24, 2017, <https://www.edweek.org/ew/articles/2017/05/24/market-is-booming-for-digital-formative-assessments.html>.
- 34 Liana Heitin, "Arizona Leans in on Formative Assessment," *Education Week*, November 9, 2015, <https://www.edweek.org/ew/articles/2015/11/11/arizona-leans-in-on-formative-assessment.html>.
- 35 "Interim Assessments," Smarter Balanced Assessment Consortium, <http://www.smarterbalanced.org/assessments/interim-assessments/>.
- 36 Jessica Bies, "How Downstate Delaware Schools Saw Biggest Jump in State Test Scores," *Delaware News Journal*, August 6, 2018, <https://www.delawareonline.com/story/news/education/2018/08/06/delaware-test-scores-education-laurel/875778002/>.
- 37 Molnar, "Market is Booming for Digital Formative Assessments."
- 38 "Formative Assessment: State Resources," The Center on Standards & Assessment Implementation, <https://www.csai-online.org/spotlight/formative-assessment-state-resources>.

- 39 "Formative Assessment Bi-Regional Advisory Board Leadership Summit," University of Oklahoma, March 27, 2018, <https://www.c3ta.org/news/formative-assessment-bi-regional-advisory-board-leadership-summit-153.html>.
- 40 "Formative Assessment for Michigan Educators (FAME)," <https://famemichigan.org/>.
- 41 Kimberly Young, "Implementation of the Formative Assessment Process in Michigan," Michigan Department of Education, 2018, <https://famemichigan.org/wp-content/uploads/2018/05/About-FAME-and-FAME-professional-learning-model.pdf>
- 42 FAME, "Research and Development," Michigan Department of Education, <https://famemichigan.org/research-and-development/#working-on>.
- 43 "Race to the Top Assessment Program: Guidance and Frequently Asked Questions," US Department of Education, <https://www2.ed.gov/programs/racetothetop-assessment/faq.html>
- 44 Nat Malkus, "A Right Turn on Assessments: State-Directed Assessments Using an Interstate Test-Item Bank Cooperative," American Enterprise Institute, December 2016, <http://www.aei.org/wp-content/uploads/2016/12/A-Right-Turn-on-Assessments.pdf>.
- 45 Ibid.
- 46 Marta Aldrich, "Why Tennessee Legislators Share Blame, Too, for TNReady Testing Headaches," Chalkbeat, April 23, 2018, <https://chalkbeat.org/posts/tn/2018/04/23/why-tennessee-legislators-share-blame-too-for-tnready-testing-headaches/>.
- 47 New Meridian, "Next Generation Assessment Options for New Jersey," New Jersey State Board of Education, August 1, 2018, <https://www.nj.gov/education/sboe/meetings/2018/August/public/Presentation%20by%20New%20Meridian.pdf>.
- 48 New Meridian, "Our Approach," Accessed March 2019, <https://newmeridiancorp.org/approach/>.
- 49 Ibid.
- 50 Email exchange, Kristopher John, New Meridian, April 2019.
- 51 "Spring 2015 Michigan Statewide Assessment Transition," Michigan Department of Education, 2015, [https://www.michigan.gov/documents/mde/Assessment\\_Transition\\_FINAL\\_11-13-14\\_473989\\_7.pdf](https://www.michigan.gov/documents/mde/Assessment_Transition_FINAL_11-13-14_473989_7.pdf).
- 52 Ibid.
- 53 Linda Darling-Hammond, "Developing and Measuring Higher Order Skills: Models for State Performance Assessment Systems," Learning Policy Institute, March 17, 2017, <https://learningpolicyinstitute.org/product/models-state-performance-assessment-systems-report>.
- 54 Every Student Succeeds Act, Public Law 114-95, §1111(a)(1)(B)(v)(ii).
- 55 Jen Kalaidis, "Bring Back Social Studies: The Amount of Time Public-School Kids Spend Learning About Government and Civics Is Shrinking," The Atlantic, September 23, 2013, <https://www.theatlantic.com/education/archive/2013/09/bring-back-social-studies/279891/>.
- 56 Ibid.
- 57 "Transforming Science Assessment: Challenges and Recommendations for States," Achieve, June 2018, <https://www.achieve.org/files/Science%20Assessment%20Challenges%20062718.pdf>.
- 58 National Research Council, *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas* (Washington, DC: The National Academies Press, 2012), <https://doi.org/10.17226/13165>.
- 59 "Why K-12 Science Standards Matter — And Why the Time Is Right to Develop Next Generation Science Standards," National Science Teachers Association, <https://ngss.nsta.org/why-standards-matter.aspx>.
- 60 Ibid.
- 61 "Using Phenomena in NGSS-Designed Lessons and Units," Achieve, September 2016, <https://www.nextgenscience.org/sites/default/files/Using%20Phenomena%20in%20NGSS.pdf>.
- 62 Disciplinary Core Ideas: Major ideas in K-12 science curriculum. These fall into four domains: the physical sciences, the life sciences, the earth and space sciences, and engineering, technology and applications of science.

- 63 Cross Cutting Concepts: Ideas that have application across all domains of science. Examples include patterns, similarity and diversity; cause and effect; stability and change.
- 64 Science & Engineering Practices: Practices that describe the behaviors scientists use to investigate and make sense of the world.
- 65 "3-ESS2-2 Earth's Systems," Next Generation Science Standards, <https://www.nextgenscience.org/pe/3-ess2-2-earths-systems>.
- 66 US Department of Education, "Applications for New Awards: Competitive Grants for State Assessments Program."
- 67 Jodi Davenport, "Toward a Transformation of Science Education: Simulation-Based Formative Assessment," WestEd, February 23, 2018, [https://wested.org/rd\\_alert\\_online/toward-transformation-science-education-simulation-based-formative-assessment/](https://wested.org/rd_alert_online/toward-transformation-science-education-simulation-based-formative-assessment/).
- 68 Interview, Rebecca Kantar, Imbellus.
- 69 "Science Assessment System through Course Task: Airbag Effectiveness," Kentucky Department of Education, [https://education.ky.gov/curriculum/conpro/science/Documents/Airbag\\_Effectiveness\\_TCT.pdf](https://education.ky.gov/curriculum/conpro/science/Documents/Airbag_Effectiveness_TCT.pdf).
- 70 "State Spotlight: Kentucky's State-Led Science Assessment System," Achieve, October 2018, [https://www.achieve.org/files/Assessment%20Systems%20Brief\\_Kentucky\\_100418.pdf](https://www.achieve.org/files/Assessment%20Systems%20Brief_Kentucky_100418.pdf).
- 71 OpenSciEd, <https://www.openscienced.org/openscienced/>.
- 72 "50 State Comparison: Social Studies Assessments Grades (if Any)," Education Commission of the States, April 2018, <http://ecs.force.com/mbdata/mbquestrt?rep=SUM1804>.
- 73 Jan Brennan and Hunter Railey, "The Civics Education Initiative 2015–17," Education Commission of the States, September 2017, <https://www.ecs.org/wp-content/uploads/The-Civics-Education-Initiative-2015-2017.pdf>.
- 74 Dafna Linzer, "The Problem With Question 36: Why Are So Many of the Answers on the US Citizenship Test Wrong?" Slate, February 23, 2011, <https://slate.com/news-and-politics/2011/02/u-s-citizenship-test-why-so-many-of-the-answers-are-wrong.html>.
- 75 Office of Superintendent of Public Instruction, "OSPI-Developed Social Studies Assessments," Washington Department of Education, <http://k12.wa.us/SocialStudies/WhatAreCBAs.aspx>.
- 76 Ibid.

## Acknowledgments

The authors would like to thank the people who shared their time, knowledge, and expertise to inform this paper, including assessment experts and professionals, current and former state education agency staff, and subject-matter experts. Special thanks to our Bellwether colleagues Andy Rotherham, Gwen Baker, and Jennifer O’Neal Schiess, who offered valuable feedback. Thank you to the Charles and Lynn Schusterman Family Foundation for their support of this work.

The following organizations mentioned in this brief are or were clients of funders of Bellwether Education Partners: the Council of Chief State School Officers, the Louisiana Department of Education, the PARCC consortium, WestEd, and New Meridian. The authors maintained editorial control of this publication. A full list of Bellwether’s clients and funders is available at [BellwetherEducation.org](http://BellwetherEducation.org).

Many people played a part in this work, but the conclusions and content of this paper are the responsibility of the authors alone, as are any errors of fact or omission.

## About the Authors



### Bonnie O'Keefe

Bonnie O'Keefe is an associate partner on the Policy and Evaluation team at Bellwether Education Partners. She can be reached at [bonnie.okeefe@bellwethereducation.org](mailto:bonnie.okeefe@bellwethereducation.org).



### Brandon Lewis

Brandon Lewis is an analyst on the Policy and Evaluation team at Bellwether Education Partners. He can be reached at [brandon.lewis@bellwethereducation.org](mailto:brandon.lewis@bellwethereducation.org).



## About Bellwether Education Partners

Bellwether Education Partners is a national nonprofit focused on dramatically changing education and life outcomes for underserved children. We do this by helping education organizations accelerate their impact and by working to improve policy and practice.

Bellwether envisions a world in which race, ethnicity, and income no longer predict opportunities for students, and the American education system affords all individuals the ability to determine their own path and lead a productive and fulfilling life.



© 2019 Bellwether Education Partners



This report carries a Creative Commons license, which permits noncommercial re-use of content when proper attribution is provided. This means you are free to copy, display and distribute this work, or include content from this report in derivative works, under the following conditions:



**Attribution.** You must clearly attribute the work to Bellwether Education Partners, and provide a link back to the publication at <http://bellwethereducation.org/>.



**Noncommercial.** You may not use this work for commercial purposes without explicit prior permission from Bellwether Education Partners.



**Share Alike.** If you alter, transform, or build upon this work, you may distribute the resulting work only under a license identical to this one.

For the full legal code of this Creative Commons license, please visit [www.creativecommons.org](http://www.creativecommons.org). If you have any questions about citing or reusing Bellwether Education Partners content, please contact us.