



## Designing Common State Assessment Systems April 2010

Earlier this month, the National Governors Association Center for Best Practices (NGA Center) and the Council of Chief State School Officers (CCSSO) released draft Common Core State Standards, the result of a state-led collaborative process to develop common standards in English-language arts and mathematics. **The nation's governors and chief state school officers believe these new standards offer an unprecedented opportunity for states to work together to dramatically improve the quality, cost-effectiveness, and comparability of state assessments.** Adoption of common standards and assessments remains a state decision. During their annual Winter Meeting, governors expressed a strong preference for having only one or two summative assessments that could be administered in all states adopting the Common Core.

In addition, the U.S. Department of Education has announced a Race to the Top Assessment Program that will make up to \$350 million available for consortia of states to develop a "next generation" of higher-quality assessments. The upgrade is necessary. Not only are definitions of proficiency inconsistent from state to state, but also today's tests are not always well-aligned with standards and tend to be focused on low-level multiple-choice questions that cannot measure the full breadth and depth of learning the Common Core State Standards expect. Moreover, current tests were not designed to provide a rich variety of timely information useful to decision-makers at all levels, including classroom teachers.

States have an historic opportunity to use Race to the Top funds to create next-generation assessment systems that can better fulfill the many purposes we have for assessment, providing rich summative data that can inform decision-making while also informing and inspiring high-quality instruction in classrooms. The next generation of state assessments can make the Common Core State Standards concrete and meaningful to educators, students, and parents and provide a critical vehicle for ensuring that all students master essential knowledge and skills.

Recognizing that opportunity, CCSSO and the NGA Center convened in February for a series of conversations with leaders of the six overlapping state consortia that already had formed to seek Race to the Top funds. Participants explored key priorities driving each consortium and identified areas of agreement that would provide a basis for common action. Those conversations yielded important agreements that will greatly facilitate collaboration to improve the quality, cost-effectiveness, and comparability of state assessments. Leaders of the consortia:

- embraced a common vision for assessment;
- developed a list of shared priorities for leveraging Race to the Top funds to design next-generation assessment systems;
- merged their efforts considerably to reduce the number of consortia moving forward; and
- agreed to participate in a joint NGA-CCSSO project to ensure comparability of summative assessment results across consortia and to reduce costs by collaborating on other activities.

## A Common Vision

Participants reached a strong consensus on a powerful vision for next-generation assessments, based on the following overarching principles:

1. *Assessments should be fully aligned with the new Common Core State Standards and measure the full breadth and depth of knowledge and skills described in those standards.*
2. *Assessments should produce a range of sophisticated data necessary to support decision-making at all levels, including indicators as to whether students are ready or “on track” to be ready for college and careers; measures of student growth over time in addition to annual performance against standards; and information on how students perform compared with their peers internationally.*
3. *States must create coherent assessment systems comprised of multiple integrated components, including a variety of formative assessments that inform, support, and improve classroom instruction, rather than continuing to rely on one annual test to accomplish too many purposes.*

## Shared Priorities

Moving beyond that broad vision, participants also agreed on a more detailed set of “shared priorities” for the design and development of next-generation assessment systems, including the following:

- Leverage cross-state collaboration to ensure comparability of summative assessment results and to promote cost-efficiency by exploiting economies of scale for research, development, and administration.
- Employ a robust mix of test questions and performance assessments\* necessary to measure the full breadth and depth of the Common Core State Standards. Although decisions about item types should be based on an analysis of how best to measure the standards, new summative assessments will likely need to incorporate a larger proportion of more sophisticated multiple-choice questions, constructed-response (or “fill-in-the-space”) questions, on-demand performance tasks, and—to the extent feasible—classroom-based performance assessments.
- Aggressively pursue technology-based solutions for more efficient delivery and scoring of state assessments and to report test results more rapidly, clearly, and in various formats that are useful both for accountability and for improving classroom instruction.
- Employ “universal design principles,” strategies for developing new assessments in ways that allow the widest possible range of students to participate fully from outset, along with appropriate accommodations to ensure maximum participation of students with special educational needs.

---

\* Performance assessments are ways to measure students’ knowledge and skills that go beyond asking them to answer multiple-choice or fill-in-the-space questions. Typically, students are asked to complete a hands-on task that can take 40 minutes or can be completed over several class periods. For example, students might be asked to research and write a magazine article or to conduct and explain the results of a scientific experiment.

- Support and involve classroom teachers in efforts to improve assessment at all levels by:
  - Providing teachers support materials and tools (curriculum frameworks, syllabi, banks of curriculum-embedded performance tasks) which help teachers to become “assessment literate”; and
  - Offering teachers opportunities to participate in the development and implementation of new state assessments, including the design of constructed-response items and performance tasks for summative assessments.
- Ensure that the design process works for states rather than against them. States should own the processes and products of assessment development. Consortia should negotiate streamlined peer review of state assessment programs by the federal government so that peer reviews are conducted at the *consortium level* rather than for each individual state.

### **Similarities and Differences between the Consortia**

Based on the consensus reached on the vision and priorities, the original six consortia agreed to substantially merge their efforts, resulting in only two major assessment consortia moving forward. Although the two consortia share a similar set of long-term priorities for the next generation of state assessment systems, they have chosen to emphasize a different set of priorities over the short term, during which Race to the Top funds will support major investments in research and development. (The appendix includes detailed descriptions provided by leaders of each consortium and its design plans as of March 2010.)

The **SMARTER BALANCED** consortium plans to move very aggressively toward full implementation of online testing using “computer-adaptive” software that selects new test questions based on each student’s own in-test performance and provides immediate results to teachers. This approach allows for a very precise understanding of where students are in relation to grade level expectations. In addition to summative tests, the consortium will develop computer-adaptive mid-year “benchmark” tests and formative assessments that can be administered throughout the year to guide instruction. The consortium plans to seek additional funds in addition to a Race to the Top grant so that it can meet these ambitious goals. Finally, SMARTER BALANCED plans to emphasize teacher involvement in all aspects of assessment design and implementation, from creating and selecting test questions to scoring students’ responses on performance tasks, and to invest heavily in other professional development activities so teachers become more adept at applying a range of assessment strategies in their own classrooms.

Recognizing the fiscal costs of research and development plus ongoing administration, the **Partnership for Assessment of Readiness for College and Career** has chosen to adopt a prudent set of short-term design goals. The Partnership will focus primarily on developing a set of high-quality summative assessments, including grades 3-8 tests and end-of-course high school tests, which can provide rich information on students’ annual progress toward meeting evidence-based benchmarks for college- and career-readiness. The intent is to ensure high-quality data that can be used for a range of planning and accountability purposes. For example, several leaders of the Partnership have a particular interest in using the results to generate student growth measures that can be used to evaluate teacher and principal effectiveness. To help teachers and others better

understand and plan for the assessments, the Partnership will release a significant proportion of test items and interpretive information every year, and it will develop model curriculum frameworks and course syllabi that illustrate specific instructional options.

Like the SMARTER BALANCED consortium, the Partnership plans to develop a bank of classroom-based performance tasks, the results of which eventually could be incorporated into summative scores, and it intends to move to computer-based testing by 2016. However, the Partnership currently does not plan to invest in computer-adaptive software, extensive teacher development activities, or mid-year benchmark and formative assessments (though its system architecture will allow states to add such components later).

### **Moving Forward: Ensuring Comparability in Summative Assessments**

CCSSO and the NGA Center will lead a joint effort to enable test scores to be compared across the summative assessments being created by the two consortia. Governors and chief state school officers view this as a top-priority goal and consider the effort to be the natural successor to the state-led effort to develop Common Core State Standards. High-quality assessments make standards concrete and meaningful to educators, students, and parents and provide a critical vehicle for ensuring that all students master important knowledge and skills. The assessment work will engage any consortium and all states that adopt the Common Core State Standards.

In the April 6<sup>th</sup> notice for applications, the U.S. Department of Education announced two competitions. A Comprehensive Assessment System grant would be for grades 3-8 and at a designated point in high school, and would be used for instructional improvement and accountability. They anticipate funding 1-2 consortia at approximately \$160 million each, with the requirement that at least 15 states be in a consortium. An additional \$30 million will fund a second grant for High School Course Assessments, which would be used to create more consistent levels of rigor in high school courses. In advance of submitting their assessment proposals to the U.S. Department of Education on June 23<sup>rd</sup>, CCSSO and the NGA Center will ask all consortia to sign a memorandum of understanding that commits them to participating in the effort to compare scores across assessments.

The NGA Center and CCSSO will engage testing experts who can help participants understand the many possible methods for promoting comparability, as well as the tradeoffs among different strategies. Certain approaches, such as embedding a set of common questions across the tests, would enable results to be compared among schools and districts across most, if not all, of the nation, something that never before has been possible in the United States. For example, even if West Virginia and Florida joined separate assessment consortia, it would be possible to say “Byrd Elementary School’s 85 in West Virginia is higher than Hurston Elementary School’s 80 in Florida.” Other approaches would require a smaller up-front investment but would not produce reliably comparable scores at the school level or district level.

In addition to ensuring a level of comparability across summative assessments, the effort will encourage consortia to explore other ways they can pool their efforts and leverage greater economies of scale. Additional cost savings might come from cross-state collaboration to develop curriculum-embedded tasks and materials based on the Common Core State Standards, or ways to evaluate and meet states’ differing technological infrastructure needs.

The full power of the Common Core State Standards will be realized when states align them to new, high-quality assessments that are internationally benchmarked and build toward college and career readiness. However, the discussion about developing common state assessments is

relatively new and lacks agreement on the one best design. While a single testing consortium would allow states and families to compare achievement at the student level, a capability that is lost with multiple consortia, there are philosophical and practical differences across states that make a single consortium difficult to create at this time. The overarching goal is to create next-generation assessment systems that are more comparable across the nation and more cost-efficient than ever before.

## Appendix: Detailed Descriptions Provided by Assessment Consortia

The following descriptions were provided by leaders of the two assessment consortia.

### The SMARTER BALANCED Consortium

#### Overarching Vision and Goals

The “Smarter Balanced Assessment Consortium” was formed from a merger of three consortia that emerged in January 2010 in response to the Race to the Top competition: the Balanced Assessment, MOSAIC, and SMARTER Consortia, comprising a total of 45 states.

The Consortium’s priorities for a new generation assessment system are rooted in a concern for the valid, reliable, and fair assessment of the deep disciplinary understanding and higher-order thinking skills that are increasingly demanded by a knowledge-based economy. These priorities are also rooted in a belief that assessment must support ongoing improvements in instruction and learning, and must be useful for all members of the educational enterprise: students, parents, teachers, school administrators, members of the public, and policymakers.

The Consortium recognizes the need for a system of formative and summative assessments, organized around the Common Core State Standards, that support high-quality learning and the demands of accountability, and that balance concerns for innovative assessment with the need for a fiscally sustainable system that is feasible to implement. The efforts of the Consortium will be organized to accomplish these goals.

The Consortium is committed to the development of a system that is state-led and will provide:

- **Common summative tests in English language arts and mathematics** that assess student progress and mastery of core concepts and critical transferable skills using a range of formats: selected-response and constructed-response items, and performance tasks, designed together to assess the full range of standards.
- **Formative assessment tools and supports** that are shaped around curriculum guidance which includes learning progressions, and that link evidence of student competencies to the summative system.
- Focused **professional development** around curriculum and lesson development, as well as scoring and examination of student work
- **Reporting systems** that provide first-hand evidence of student performances, as well as aggregated scores by dimensions of learning, student characteristics, classrooms, schools, and districts.
- A **governance structure** that ensures a strong voice for state administrators, policy makers, school practitioners, and technical advisors to ensure an optimum balance of assessment quality, efficiency, costs, and time.

## Design Principles

As described below, the Consortium members have agreed to a set of principles that are consistent with those used by educational systems of high-achieving nations and states. These include the following:

1) **Assessments are grounded in a thoughtfully integrated learning system** of standards, curriculum, assessment, instruction, and teacher development. Teachers and other instructional experts are involved in the process of developing formative and summative assessments grounded in the learning standards. These guide professional learning about curriculum, teaching, and assessment. Instructional supports are provided to enable thoughtful teaching. Thus, assessments are provided to schools as part of a well-aligned system that guides and supports a coherent approach to students' and teachers' learning.

2) **Assessments include evidence of actual student performance** on challenging tasks that evaluate standards of 21<sup>st</sup> Century learning. The assessments will be strategically used to evaluate a broad array of skills and competencies and inform progress toward and acquisition of readiness for higher education and multiple work domains. They emphasize deep knowledge of core concepts within and across the disciplines, problem solving, analysis, synthesis, and critical thinking.

3) **Teachers are integrally involved in the design, development and scoring of assessment items and tasks.** Teachers will participate in the alignment and unpacking of the Common Core State Standards and the identification of the standards in the local curriculum. The Consortium will involve teachers in formative and summative assessment development. It will support moderation of scoring processes to ensure consistency, to enable teachers to deeply understand the standards, and to develop stronger curriculum, instruction, and classroom assessment. Assessment literate teachers 1) who have gotten "inside" the Common Core State Standards, 2) who have taught to the standards, 3) who have learned how to appropriately measure the standards, and 4) who have learned strategies to intervene if students have not measured the standards, will be teachers whose students are learning. Teachers' roles include the construction and review of items/tasks, the definition of scoring guides, selection of student work exemplars, and scoring.

4) **Technology is designed to support assessment and learning systems.** Technology is used to enhance these assessments in a number of ways by: delivering the assessments; enabling adaptive technologies to better measure student abilities across the full spectrum of student performance and evaluate growth in learning; supporting online simulation tasks that test higher-order abilities, allowing students to search for information or manipulate variables, and tracking information about the students' problem-solving processes; and, in some cases, scoring the results or delivering the responses to trained scorers / teachers to access from an electronic platform. Such a platform can support training and calibration of scorers and moderation of scores, as well as the efficient aggregation of results in ways that support reporting and research about the responses.

5) Assessments are structured to continuously **improve teaching and learning**. Assessment, *as*, *of*, and *for* learning, is designed to develop understanding of what learning standards are, what high-quality work looks like, and what is needed for student learning. It is also designed to foster instruction that supports transferable knowledge and skills. These outcomes are enabled by several features of the assessment system:

- The use of school-based, curriculum-embedded assessments provides teachers with models of good curriculum and assessment practice, enhances curriculum equity within and across schools, and allows teachers to see and evaluate student learning in ways that can feed back into instructional and curriculum decisions.
- Close examination of student work and moderated teacher scoring are sources of ongoing professional development that improve teaching.
- Developing both on-demand and curriculum-embedded assessments around learning progressions allows teachers to see where students are on multiple dimensions of learning and to strategically support their progress.

### **Anticipated Uses of the Assessment System**

The Consortium will develop a common summative assessment that will provide comparable results across all of the participating states. This comparability will be achieved by applying psychometrically sound scaling and equating procedures to items and a modest number of performance tasks of limited scope (e.g. no more than a few days to complete) that will be used in common across consortium states. Consortium states will use commonly determined performance standards that are internationally benchmarked.

In addition, some states will work on pushing the envelope with respect to more ambitious performance assessments – which may be used in common by one or more sub-consortia of states – and, in the same way, others will undertake more ambitious work with respect to computer-adaptive testing and simulations. This design allows the Consortium to create, at one time, a new summative assessment used by a large number of states within the five-year horizon of the federal grant, and to create even more leading-edge assessment components used by sub-consortia of states who decide to offer augmented assessments. Common use of these augmented assessments across subsets of states would result in comparable results across the states, without disrupting the existence of a leaner, common summative assessment across all the states in the Consortium.

Current understandings about the nature of the assessment items, tasks, and strategies are noted below:

#### *Objective Machine-Scored Items*

- Movement toward more analytic types of selected-response and constructed-response items that are easily scored, including computer simulations.

#### *Open-Ended Constructed Response*

##### *Artificial intelligence (AI) scored items.*

- Work to establish efficient means of developing items and reliable scoring processes for complex responses scored by computer.
- Build and maintain the confidence teachers have in the system by incorporating a systematic read-behind by teachers.

##### *Human scored constructed response*

- Develop training and moderated scoring processes for teacher scoring of items that cannot be scored by AI and for additional scoring of AI items.
- A strategic mix of teacher and machine scoring should be created to take advantage of efficiencies and reduce burden, while also ensuring teacher participation and learning.

#### *Curriculum-Embedded Performance Assessments*

- The common summative assessment would incorporate performance events of modest scope (1-5 days) to evaluate the standards more fully.
- Some states will form a workgroup to go further with rich performance tasks that can make advances in performance assessments on behalf of the Consortium.
- These more ambitious performance assessments could be included for individual state accountability systems (and for comparisons across a subset of states, if desired) until a greater proportion of states has capacity for implementation.

#### *Advanced Computer-Based Simulations*

- Some states will form a workgroup to make advances in computer based simulations on behalf of the Consortium.
- These simulations could be included in individual state accountability systems until a greater proportion of states has capacity for implementation.

## **The Partnership for Assessment of Readiness for College and Career**

*In January 2010, twenty-eight states signed an agreement to participate in the Common Assessment Partnership and seventeen states signed with the Florida-led Common Assessment Consortium.<sup>†</sup> Since then, many leaders and assessment experts from these states have engaged in work, facilitated by Achieve, to develop a shared vision and set of design principles for a multi-state assessment system. During this period, leading states in both consortia—Florida, Massachusetts and Louisiana—worked to align the visions of the two consortia. This document represents their collective vision and a summary of current agreements and understandings.*

The Race to the Top Assessment Competition presents states with an unprecedented opportunity to move from the state-led development of common standards in mathematics and English language arts to a common measurement for student performance and growth. The Common Core State Standards will require students to demonstrate knowledge and skills in deep and meaningful ways, as well as to reason, synthesize, think critically, and solve problems. A compelling vision for common assessments demands fully measuring the depth and breadth of the concepts and skills represented in the Common Core State Standards. However, states recognize the tension between their desire for innovative, forward-looking assessments and the realities of limited resources available to them for ongoing test administration. States in this partnership have agreed to strike a balance between pushing ahead towards next-generation assessment systems and acknowledging the design and fiscal tradeoffs, including the ability to sustain these assessments over the long term.

The state leaders recognize that trying to project costs more than five years in the future is filled with many uncertainties, such as the potential cost savings from technologies that have not yet been invented. Therefore, these state leaders have agreed that they will be adamant about researching and designing an affordable and practical system without sacrificing innovative assessments that can drive instruction. Partnership states will bring forth the best intellectual resources to tackle this challenge and develop solutions that will allow states to maximize the value of innovative assessment features while minimizing cost and turnaround time for results.

In the near term, the partnership expects that the substantial costs for developing the assessment system outlined below will be paid for by the Race to the Top assessment grant award. However, the partnership members recognize that the costs of implementing and sustaining an innovative assessment system could require more resources than many states are currently budgeting for assessment, even with new technological developments. The states are committed to building a sustainable system and it is their hope that the federal government will continue to provide funding to help support the ongoing administration costs for innovative assessment systems.

### **Purposes and Uses**

The initial state members have identified the following major purposes and uses for the assessment system results.

- The primary purpose is to measure and document students' **college and career readiness** at the end of high school and to measure students' progress toward this target throughout

---

<sup>†</sup> The combined list of states: Alabama, Arizona, Arkansas, California, Delaware, District of Columbia, Florida, Georgia, Hawaii, Illinois, Indiana, Kentucky, Louisiana, Maryland, Massachusetts, Michigan, Minnesota, Mississippi, New Hampshire, New Jersey, New Mexico, New York, North Carolina, Ohio, Oklahoma, Pennsylvania, Rhode Island, South Carolina, Tennessee, and Wisconsin.

the rest of the system. Students meeting the college- and career-readiness standards will be eligible to be placed into credit-bearing, rather than remedial, courses in all public 2- and 4-year postsecondary institutions in all participating states.

- Additionally, the partnership is committed to ensuring that the assessment results:
  - Are **comparable across states** at the student level;
  - Meet **internationally rigorous benchmarks**;
  - Support valid assessment of **student longitudinal growth**; and
  - Serve as a **signal for good instructional practices**.
  
- The results must be able to support multiple levels and forms of accountability including:
  - Decisions about **promotion and graduation for individual students**,
  - **Teacher and leader evaluations**, and
  - **School accountability** determinations.

### Cross-Cutting Design Considerations

While there are many design issues unique to either the grades 3-8 or high school assessment systems, the following issues cut across the design of all of the assessments in the system.

- ***Comprehensive and Coherent System.*** A comprehensive assessment system design will be used to ensure coherence among summative, interim, and formative assessments, even if the partnership focuses development efforts on the summative measures.
- ***Operational Use.*** The partnership's summative assessment system will be available for the first operational use by the spring of 2014.
- ***Migration to Computer-Based Testing.*** The initial operational assessment will be available in both computer and paper formats, but by the spring of 2016, paper formats will be available for specific testing accommodations only.
- ***Common Performance Levels.*** All partnership states will use common performance level descriptors and standard-setting processes, and will cut scores to define common achievement levels.
- ***Student-Level Growth.*** The summative assessments will provide valid inferences regarding individual student growth and progress toward college and career readiness. Partnership members are committed to exploring the use of a common student growth model in order to facilitate comparisons of growth across member states.
- ***International Benchmarking.*** The assessments will be designed to ensure that students are being held to internationally competitive expectations via:
  - tight alignment with the internationally-benchmarked Common Core State Standards;
  - benchmarking the actual assessments against assessments from high-performing countries; and
  - pursuing empirically-based international comparisons at target grade levels.
- ***Item Types.*** The partnership will ensure that the assessments measure the depth and breadth of the Common Core State Standards and signal effective instruction. In consideration of cost, scoring time, and test administration time, the partnership will pursue innovations in item types that require higher-order thinking skills but that can be

scored via computer. There is also recognition that a target of college and career readiness requires expectations for complex performances. As such, assessments will include open-response tasks.

- **Testing Conditions.** The partnership is committed to using the most uniform test administration policies and practices possible to enable meaningful comparisons of results across states.
- **Special Populations.** The assessments will be as inclusive as possible, particularly for students with disabilities and English language learners. The partnership will also require—to the fullest extent possible—the use of uniform accommodation policies and practices in all member states.
- **Robust Writing Assessments.** The partnership will create robust (i.e., not just single prompts) direct writing assessments for every grade 3-11. All states will administer these at key grades and will be free to administer them (or allow Local Education Authorities to do so) at the other grades.
- **Classroom-Embedded Performance Tasks.** The partnership will develop classroom-embedded performance tasks, starting first with writing as described above. Partnership states will participate in a pilot administration of these embedded tasks. The results from these tasks will not be included in summative judgments until the validity of such judgments can be assured.
- **Released Items and Item Analysis.** The partnership will release operational items along with relevant student performance information (e.g., released-item reports).
- **Model Instructional Supports.** The partnership will develop model curriculum frameworks in grades K-8 and model course syllabi for high school that illustrate specific instructional options for educators targeting the Common Core State Standards, the common assessments, and embedded performance tasks.
- **Assessments in Grades K-2.** The partnership is interested in collaborating on some form of a K-2 assessment system.

## Assessment Design Considerations

**Grades 3-8.** The assessment system for grades 3-8 will provide students, parents, and educators with clear signals about whether students are on track to acquire the knowledge and skills foundational for success in and after high school. These assessments will include the following unique design considerations, as well as the cross-cutting features described above:

- Reading and mathematics assessments will be administered at the end of each school year in all grades.
- Writing will be assessed separately at specific, as yet to be determined, grades.

**High School.** The major focus of the high school assessment system will be to determine whether students can demonstrate the knowledge and skills necessary for success in college and careers.

- The partnership states are committed to involving higher education in the design of the assessments and associated performance standards.
- The partnership is committed to developing at least two approaches to high school assessment.

- **End-of-course exams** will be developed for a limited set of mathematics and English courses.
- **End-of-domain assessments** will be created to assess students at key points during their high school experience.
- The partnership is committed to designing these two approaches such that college/career ready determinations from each have comparable meanings.

## Governance

The partnership will employ a multi-level governance and management structure designed to guide the partnership through the submission of the proposal.

- The **Governing Board** will be comprised of a representative group of leaders from partnership states and will be responsible for major policy decisions such as the overall direction of the partnership, major purposes and uses of the assessment system, fiscal authority, and rules for state engagement.
- The **Design Group** will include officials from 8-12 states with expertise in assessment design and development and will work with an advisory group of national and international experts to create the design for a next-generation assessment system.
- The **Review Team** will be comprised of state representatives from all partnership states and will be responsible for providing input to and feedback on the assessment system design.
- Achieve will serve as the coordinating **management partner** with the National Center for the Improvement of Educational Assessment (Center for Assessment) serving as a technical support partner.