

## DEVELOPING AN INTERNATIONALLY COMPARABLE BALANCED ASSESSMENT SYSTEM THAT SUPPORTS HIGH-QUALITY LEARNING

This policy brief is based on a paper presented by Linda Darling-Hammond and Ray Pecheone, Stanford University, with Ann Jacquith, Susan Schultz, Leah Walker, and Ruth Chung Wei, at the National Conference on Next Generation K – 12 Assessment Systems, March 2010. Download a copy of the final paper by Darling-Hammond and Pecheone, as well as other papers presented at the conference, at <http://www.k12center.org/publications.html>.

*This model proposes a balanced assessment system that integrates curriculum and assessments, both formative and summative, and is designed to support higher-quality, more coherent instruction. It would be implemented by consortia of states committed to developing curriculum frameworks based on the Common Core Standards and learning progressions. The components of the assessment system—all designed to go beyond recall of facts and show students' abilities to evaluate evidence, problem solve, and understand contexts—include analytic selected-response items, short and extended constructed-response items, and standardized performance tasks in each grade level tested. The results of on-demand tests (which include the first two types of items) combined with weighted results of reliably scored curriculum-embedded performance tasks would provide student achievement data that could be compared across schools, districts, and states (summative assessments). The system also would be able to show student growth along multiple dimensions. Technology would be used to deliver tests, collect student work for teacher/scorer use and for training scorers, manage the scoring/reporting processes, and eventually help teachers to manage classroom practice. Technology would also support computer adaptive testing and computer scoring of some open-ended items. This system, properly implemented, would provide a more performance-based and useful assessment system at potentially no more cost than the present less rich and less helpful systems in use.*

Despite standards-based reform efforts in the United States over the past two decades, student achievement in this country continues to slide backwards compared to that of other students on international tests. The Common Core Standards under development by the states could help align expectations for our students with those of high-achieving countries, but how the topics are taught and assessed also must be addressed. Exams in the United States tend to rely on multiple-choice items that emphasize recall of discrete facts contrasts with the exams in higher-performing countries, which ask students to analyze, apply knowledge, and write extensively. The exams in other countries also are becoming more performance-based, with greater use of school-based tasks.



Center for  
K – 12 Assessment  
& Performance Management

*Created by Educational Testing Service (ETS) to forward a larger social mission, the Center for K – 12 Assessment & Performance Management has been given the directive to serve as an independent catalyst and resource for the improvement of measurement and data systems to enhance student achievement.*

## **Principles of the Model**

The priorities of this assessment system model focus on assessing deep disciplinary understanding and higher-order thinking skills, as well as on using assessment to continually improve instruction and learning. The model shares a set of principles with high-performing nations and states in this country that includes:

- Assessments are grounded in a thoughtful, standards-based curriculum and are managed as part of a tightly integrated system of standards, curriculum, assessment, instruction, and teacher development.
- Assessments include evidence of actual student performance on challenging tasks that evaluate standards of 21<sup>st</sup> century learning.
- Teachers are integrally involved in the development of curriculum and the development and scoring of assessments that include both the on-demand portion of state or national examinations and local tasks that feed into examination scores and course grades.
- Assessments are structured to continuously improve teaching and learning (e.g., school-based curriculum embedded assessments and a close examination of student work).
- Assessment systems are designed to emphasize the validity and quality of external assessment aimed at driving high-quality learning of ambitious intellectual skills.
- Assessment and accountability systems use multiple measures to evaluate students and schools.
- Assessment and accountability systems are used primarily for information and improvement.

These principles put considerable emphasis on curriculum as a key lever for two reasons. First, the curriculum expresses the kinds of learning that is sought, such as transferable learning and deep understanding of the central concepts in each discipline studied. Second, carefully constructed curriculum can help teachers understand how learning unfolds, or the typical learning progression within specific domains in a discipline.

This proposed assessment system would include: summative tests that measure student progress and mastery of core concepts with a range of assessment formats (multiple-choice, constructed-response, and performance tasks), formative assessments shaped around curriculum and learning progressions, focused professional development, and reporting systems that provide first-hand evidence of student performance along with aggregated scores.

## **Roles for Each Level of Government**

The **federal government** would continue strong investments in education, such as refining the National Assessment of Educational Progress, supporting research on curriculum and assessment design, allowing and encouraging the use of performance assessments under the Elementary and Secondary Education

Act, and funding initiatives that infuse knowledge of assessment and learning into pre- and in-service professional development.

Working through consortia, **states** would create common standards mapped across grade spans in a set of learning progressions; adopt and augment the standards as appropriate for a state's context; create and deploy a curriculum framework that incorporates the standards; build and manage an assessment system of both on-demand and curriculum-embedded assessment components that evaluate the full range of standards and allow evaluation of student progress; create an audit system for ensuring the comparability of teacher-managed or teacher-scored assessment components; and ensure that the education and development of teachers and leaders infuses knowledge of learning, curriculum, and assessment.

**Districts and schools** would use state curriculum guidance to further refine/revise and continually improve their curriculum; design, select, and incorporate formative assessments based on the standards, curriculum, and learning progressions; participate in administering and scoring relevant portions of the assessment system and examining student work and outcomes; and engage in the review of assessments and student work within and beyond the school.

## **How the Assessment System Would Work**

Drawing from successful practices already documented in this country and abroad, a consortium of states adopting this blended system would:

- Develop curriculum frameworks based on the Common Core Standards and knowledge of learning progressions, using and expanding upon evidence-based curriculum and assessment efforts.
- Create a digital curriculum and assessment library that would ultimately include materials for curriculum building, model syllabi for specific courses, and formative and summative assessment tasks and materials to train both teachers and school leaders on all aspects of the curriculum and assessments.
- Develop state and local assessments. Initially, the consortium would create a common reference examination that includes selected-response, constructed-response, and performance components for Grades 3-8 and at least one high-school grade. The curriculum-embedded components would be developed around core concepts and central skills in English language arts (ELA) and mathematics. They would link to the concepts and skills evaluated in the on-demand test and would allow the measurement of skills that require more time and student effort than current snapshot tests. Curriculum-embedded tasks used for summative purposes would be standardized, scored in moderated fashion, and the scores aggregated up to count as part of the external assessment. (Over time some of the open-ended items and tasks might also be scored, in part, by computer.) Curriculum-embedded assessments also could be used in formative assessments and give teachers useful information for instruction.

All components of the system would incorporate principles of universal design intended to remove construct-related barriers for non-native English speakers and students with other specific learning needs.

The model addresses measurement of student growth along a continuum of performance standards and related learning progressions, not limited by grade level and placed on a vertical scale, in both on-demand exams and the more extended classroom assessments.

## **At the High-School Level**

The consortium would explore several options for assessment at the high-school level, including:

- Course- or syllabus-based systems like those used in many countries and by the International Baccalaureate (IB) Programme and the New York Regents system. These would include within-course performance assessments and high-level end-of-course exams.
- Standards-driven systems with a more comprehensive benchmark assessment in ELA and mathematics, complemented by collections of evidence that demonstrate students' abilities to meet certain standards in and across the disciplines. These assessments allow more curricular flexibility in meeting the standards.
- A mixed model that includes both course- and standards-driven models, allowing some demonstration of proficiency in any given course or even outside the boundaries of a course.
- Develop moderation and auditing systems for teacher-scored work. Training would ensure consistent scoring.

Teachers and school leaders will also need time and training to implement an integrated system of curriculum, instruction, and assessment. States and districts must be committed to support teacher engagement in curriculum and assessment development, scoring, and analysis. Moreover, states will need to develop technology to support the system.

## **Technology Use and Reporting**

A technology platform is key to this assessment model and would be used to:

- Deliver both on-demand and curriculum-embedded assessments to students and teachers.
- Make use of adaptive technologies to deliver tests electronically and create assessments that are responsive to the test taker's performance.
- Deliver online tasks of higher order abilities.
- Score multiple-choice and some open-ended items and tasks.
- Deliver responses to other tasks that are assessed by trained scorers/teachers using an electronic platform.

- Support training and calibration of scorers and moderation of scores.
- Enable efficient aggregation of results in ways that support reporting and research about the responses.

This model would build an interactive web-based platform that allows robust and efficient collection, sharing, evaluation, and aggregation of evidence about student learning. Digitized student responses will allow states to manage all the relevant balanced assessment data on a common platform and give students and teachers information on assessment performance rapidly. The model permits tasks tailored to individual students, efficient delivery of training on scoring, audit scoring, and report generation.

### **Accessibility and Technical Quality**

With the guidance of a blue-ribbon panel, this assessment system model will help develop the tools, strategies, and specific modifications/accommodations that will allow a wide range of students with specific language or learning needs to be assessed with accessible items and performance tasks/items aligned to college-ready standards

### **Informing Instruction and Leadership**

The ongoing information loop that is built into this blended assessment system will provide the data for continuous learning by students, teachers, school leaders, and policymakers. Aligning the assessment system to common standards will enable cross-state and within-group analyses of student content knowledge and foster the capacity to make decisions about professional development, teacher preparation, technology, instructional interventions, and the use of school time and resources.

### **Goals and Timelines**

The development work will begin with a comprehensive review of consortium assets, best practices, and challenges. A technical advisory committee will be convened to guide the design and development of the balanced assessment system, which will be the focus of the first year of implementation, followed by small-scale pilots and any needed modifications in the second year while further design and development work is being completed. Field trials and building capacity for scale up of the assessment model follows in the third year as areas needing further pilot testing are fine-tuned. Readiness for full scale up across the consortium will occur by Year 5.

### **Costs**

A single dollar figure of test expenditures does not capture the cost-benefit trade-offs such as the benefits of instructionally useful assessments or, conversely, the narrowing of the curriculum because assessment systems choose what may appear to be the least costly item types.

This paper includes a detailed cost model that used empirical estimates of costs from current testing programs to estimate expected savings from certain kinds of efficiencies. The analysis estimates that a

contemporary largely multiple-choice test battery in a typical state costs about \$20 per student. In the same typical state, a high-quality assessment that includes more constructed-response items and an average of two to four additional performance tasks would cost about \$55 per student. However, three strategies can yield cost savings: participating in a consortium, using online technology to deliver and score tests, and using teacher-scoring of open-ended items (rather than vendor scoring) and some computer-based scoring of open-ended items. Combining all of these strategies would reduce the cost of high-quality assignments to about \$10 to \$20 per student, depending on the teacher-scoring model used. The authors contend that, with careful planning and attention to the costs of various decisions, it is possible to develop and administer the system they propose at no more than the same cost we are paying for current tests.

### **For More Information**

For more information on this assessment system model, please see the paper by Linda Darling-Hammond and Ray Pecheone:

Darling-Hammond, L., & Pecheone, R. (2010). *Developing an internationally comparable balanced assessment system that supports high-quality learning*. Retrieved from <http://www.k12center.org/publications.html>.

For more information on the National Conference on Next Generation Assessment Systems, please see: <http://www.k12center.org/events.html>.