

## AN AMERICAN EXAMINATION SYSTEM

This policy brief is based on a paper presented by Lauren Resnick, University of Pittsburgh, and Larry Berger, Wireless Generation, with Brian Junker, Carnegie Mellon University, at the National Conference on Next Generation K – 12 Assessment Systems, March 2010. Download a copy of the final paper by Berger and Resnick, as well as other papers presented at the conference, at <http://www.k12center.org/publications.html>.

*This model focuses on designing an assessment system that works to improve instruction. The model reflects key aspects of the substantive, cognitively demanding European systems, while maintaining standards of psychometric rigor necessary to support America’s accountability, comparability, and equity agendas. The model aligns standards, assessment, and curriculum through three principal innovations: distributed accountability exams that have a high degree of content and instructional validity; a system of “mass personalized” formative assessments that become an integral part of each teacher’s instructional routine; and a technology platform with broad uses, especially to help teachers manage the assessment process and have ready access to insights from the assessment data. A major aspect of the technology is the creation of a honeycomb, or interactive map, that visually explains the instruction and assessment goals in each grade as well as across grades, tracking the progress of individual students, classes, schools, and districts. A substantial amount of work has already been done to develop the content and tools needed to implement the American Examination System.*

### The Problem

For two decades the United States has been trying to move to a standards-based accountability system, but instead it has created a test-based accountability system that does not reflect the standards set at the beginning of the 1990s, much less today’s “fewer, clearer, higher” Common Core Standards. Several studies have shown that most state tests do not measure the higher order thinking, problem-solving, and creativity needed for students to succeed in the 21st century. These tests, with only a few exceptions, systematically over-represent basic skills and knowledge and under-represent the complex knowledge and reasoning we are seeking for college and career readiness. This has had negative effects on teaching and learning, especially for poor and minority students. The use of interim assessments as a check on progress tends to reinforce these suboptimal teaching behaviors.

This situation has created a testing bind that drives attention away from the intended standards. The effects are greatest in the poorest schools. And it may be lowering the learning opportunities even for many more privileged children as schools turn their energies to the test-based basic skills program.



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.*

## A Solution

Testing and accountability should remain at the heart of national education policy because equity and national prosperity depend on an education system that keeps stretching toward higher achievement, but there should be new forms of assessment. The proposed American Examination System draws from some of the most successful policies/practices overseas, while maintaining goals important to this country such as equity and comparability. This model:

- Emphasizes the kind of instruction that is valued so that preparing students for tests works for—rather than against—high-cognitive demand instruction.
- Places exams within ongoing instruction so that assessments support instruction rather than distract from it.
- Ensures content and instructional validity of all assessments, avoiding the alignment problem that has plagued state-testing programs.
- Provides reliable and valid accountability for student, school, and educator performance.
- Includes diagnostic tools for instruction to meet individual student needs.
- Leverages advanced data collection and computational resources for mass personalization of the formative assessments for students and teachers.

The American Examination System will be *educative* for those who use it—students, teachers, and education organizations—and will help them do their jobs better. The system will accomplish these goals through *distributed accountability exams* and *formative assessments* linked to specific content topics and managed through an online technology platform.

### Distributed Accountability Exams (DAEs)

Accountability data for this system would come from exams administered throughout the school year after students have completed a unit of study on particular content and skills based on the Common Core Standards. The exam tasks would be familiar to students because they would be similar to the instruction they had received, but neither teachers nor students would know the questions beforehand. Students would take three to five distributed accountability exams (DAEs) in mathematics and literacy during each school year, with each exam assessing material covered through 3–7 weeks of instruction. The specific number and timing of exams would be worked out with the states.

The DAEs would model the kind of high-cognitive demand performance embedded in the Common Core Standards, as well as test basic skills. The standards provide a foundation for a criterion-referenced exam system that is closely tied to instruction yet meets technical quality criteria. The core grade-level standards are specified at a granular size that can be used to organize meaningful units of instruction and correspondingly meaningful assessments. Tasks/items for the DAEs would go through rigorous processes, including scientific (experiment-based) tests, to establish *content* and instructional validity.

Availability of multiple forms of the DAEs (to use as pre- and post-tests) would allow states and districts to use the content-based exams to plot student growth, as well as teacher and school effectiveness.

The DAEs would contain a mix of short constructed-response items and more extended written responses (needing human scoring), along with multiple-choice items as appropriate. The exams, given several times over the course of the year, would provide much more data than current end-of-year tests, thereby increasing test reliability.

## **Educative Formative Assessments**

The model would foster the development and use of formative assessments for daily/weekly classroom use that in part mirror the summative assessments. These educative formative assessments would be aligned with the learning trajectories from the Common Core Standards and thus aligned with *what* teachers need to teach. The assessments would model approaches to *how* to teach and would be part of teachers' instructional routines rather than be imposed as an extra testing assignment. Formative assessments that cannot be machine-scored would have simple rubrics that can be quickly analyzed. Teachers' scoring forms would be digitalized and include samples of answers for each rubric so teachers can calibrate their analyses.

The formative assessment results should not be used in accountability reporting, but the student, class, and school results would be available to teachers and principals for use in classrooms and for professional development. Metrics of fidelity in implementing the formative assessments (and their associated instructional recommendations) could be used as part of teacher/school performance management/accountability. For instance, formative data can show whether teachers are doing progress monitoring with the frequency appropriate for each student. (The DC public school system is an example of a school system that is already using these types of formative assessment metrics as part of their SchoolStat approach to continuous, district-wide performance management.)

## **Managing Mass Personalization**

An advanced model of measurement can be built on a great magnitude of data—both formal and informal—on each student in the course of the year so that each test enhances a picture already drawn of the student. This mass personalization, already used in Internet-based commerce and social networking, eventually will be able to personalize each assessment at the individual student level. Attributes that could be the basis of personalization include past student performance on assessments, teacher and school characteristics, aggregated assessment performance of students in a school, previous effectiveness of the teacher, and the specific curriculum and assessments used.

Technology makes this scalable—the amount of data computers can collect is virtually unlimited. The initial goal for mass personalization would be to apply it to customizing formative assessments used as part of classroom instruction.

## The Assessment Platform

The assessment platform would manage both the DAEs and the formative assessment system, enabling assessment delivery, scoring, reporting, and analysis. Based on widespread classroom experience with existing products and on current designs (some of which have been funded by the Gates Foundation), the platform would be able to handle all of these elements at scale in a cost-effective way, while minimizing additional burdens for teachers, students, and administrators.

The model provides a *honeycomb* or an interactive map of learning trajectories. It shows the instruction and assessment that should take place across all grades, collecting data on the individual progress of students, classes of students, schools, and school districts. It also would provide data to validate/refine hypotheses about the skills in the Common Core Standards and state standards.

When the system is fully operative, educators will have formative and summative assessments for each skill step along each learning trajectory, starting with mathematics and literacy for Grades 3–10. Educators will be able to see exam and formative assessment results for each student and how the results match with the trajectory.

The assessment platform will enable students to take the assessments online or on paper, support teachers/schools in scanning and uploading paper-based assessments and other student work, manage remote scoring workflow, provide teachers with a scoring interface and dashboard tools for tracking and analyzing the progress of particular students or groups of students, provide administrators with a reporting interface that includes aggregate analysis, and allow teachers to share formative assessments with each other. The platform also will include an assignment builder, so that educators can select formative assessment items as tasks for use by the students in the classroom or as homework.

The American Examination System would not assume, at the outset, that all assessments will be conducted with students sitting at computers. Given current school infrastructures and the challenge of showing mathematics work via keyboard, it may be more efficient to continue to rely for some time on paper-and-pencil inputs to an otherwise digital system. The continued value of these “primitive” recording tools seems especially compelling when one considers that much of the value of the new generation of assessment tasks depends on soliciting open-ended expressions of student reasoning and thinking—and in the case of mathematics this includes drawings, graphs, and explanations. So the American Examination System would include a process to enable scanning/digital photographing, uploading, and archiving of very large volumes of paper-based student work, including work for DAEs, to enable remote scoring as well as online student portfolios. The scanning/photographing process, which has already been tested in North Carolina classrooms, puts minimal burdens on teachers or other school staff and does not require large per-school investments in hardware or network infrastructure.

## Development Timelines

This assessment model is feasible because much of the work in developing the content and tools needed already has been done. The system can be fully operational within 3–4 years from the beginning of the process, with mass personalization of summative assessment playing a larger role at the end of that timeframe.

## **Costs**

We estimate that a typical state would spend the same for the American Examination System as it now spends on assessments to meet No Child Left Behind requirements, or \$20-30 per student. The DAEs will cost more to administer than current high-stakes tests because they will be given more frequently. On the other hand, states will be able to eliminate the costs of current interim testing.

## **For More Information**

For more information on this assessment system model, please see the paper by Lauren Resnick and Larry Berger:

Resnick, L., & Berger, L. (2010). *An American examination system*. 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>.