Reimagining standardized testing: How a hybrid assessment system can benefit students, teachers, and administrators

By Steven Shapiro

December 2014

Key points:

- America’s K–12 standardized-testing system has fallen short in providing educators with the necessary tools to improve student achievement.
- Administering frequent hybrid exams—which combine traditional classroom-summative tests with state-standardized test questions—could help students develop their skills, help teachers adapt their instructional strategies, assist administrators in efficiently allocating resources, and give local and state leaders a better data platform upon which to base policy decisions.
- This type of integrated assessment infrastructure would be particularly useful in mathematics, where a progression of content mastery is essential for the cumulative learning process, and would better prepare students for both college and careers.

Over the past two decades, a key element of the US K–12 school reform movement has been a nationwide campaign to implement standardized-testing programs to quantify learning outcomes, improve accountability, and, ultimately, raise student achievement. Thus far, there is little evidence of significant improvement. Although more students are taking Advanced Placement (AP) and International Baccalaureate (IB) courses, less than 40 percent of high-school graduates are prepared for college, and only 59 percent receive their degrees within six years.¹

Furthermore, businesses often express concern that many job applicants lack the qualifications needed for middle-skill jobs (those that require some postsecondary training but not a four-year college degree, such as those of welder, auto mechanic, computer network administrator, or electrician), which are estimated to constitute about 45 percent of all job openings.²
The heightened emphasis on standardized tests has provoked considerable controversy. Beginning in 2001 with the passage of the No Child Left Behind Act (NCLB), many organizations expressed concerns about the quality of state standards and their associated exams, the misinterpretation of test data, and the impact of high-stakes testing on instruction, curricula, and school climate.3 “Teaching to the test” became a pejorative phrase. The Common Core State Standards (CCSS) initiative was designed to address some of these concerns, but its implementation has brought a new set of controversies.

Teachers, schools, and state agencies are making a difficult transition to new instructional strategies and more-rigorous standardized assessments. In the first half of 2014, some states abandoned the CCSS or withdrew from testing consortia, citing undue federal influence over local educational policy. Amid all the heated rhetoric, the underlying argument for standards, standardized tests, and their relationship to student achievement is getting lost in the noise. Standardized testing can play a useful role, but how it is utilized needs to be reconsidered.

This paper describes a new approach to constructing, administering, and interpreting standardized tests, envisioning an integrated system in which state-standardized exams are interwoven with the classroom-summative assessments throughout the school year. By modularizing standardized tests, the pace of instruction and assessment could become more personalized, feedback could become more timely and useful, and more granular proficiency data could facilitate a rethinking of the true meaning of “college and career readiness.”

It is important to note that I am neither a psychometrician nor an expert in assessment design. Rather, I have spent the past 13 years as a math teacher at two public high schools that serve predominantly low-income minority students, including many new immigrants and those with special needs.4 I have served as a teacher leader and participated in assessment-development programs for a district that serves almost 200,000 students. Before my teaching career, I spent more than 20 years as a business executive and management consultant for several leading technology companies.
The school from which I recently retired is fortunate to have a dedicated and competent team of administrators and teachers whose collective efforts have helped the school gain national recognition. Based on student performance on both state standardized tests and AP exams, the school was a silver medalist in *U.S. News and World Report*’s 2014 rankings of the best high schools in the country. It was also one of the highest-achieving low-income schools on the Organization for Economic Co-operation and Development (OECD) Test for Schools (based on the Program for International Student Assessment), which was piloted in 2012. Although these accolades might appear to be good news, I have also witnessed firsthand the impact of current standardized-testing policies on the emotional, logistical, and pedagogical climate of the school and have become concerned about how we are preparing the next generation for life beyond the classroom.

For the remainder of this paper, I draw on my personal experiences as a math teacher, focusing specifically on middle- and high-school math. I also limit my discussion to tests aligned with state standards that are used to measure student achievement and that comply with federal accountability mandates. Although I refer to the CCSS as a framework for explaining some concepts, I do not intend to endorse or criticize any particular set of standards, curriculum, or pedagogical strategy. Rather, I aim to illustrate how standardized testing can more effectively support the learning process and give local and state leaders the data platform upon which to base policy decisions.

With that, let’s first take a look at the logic undergirding standardized testing and where it has fallen short.

**Standardized Testing: A History of Failed Expectations**

Although most standardized tests were initially developed to sort populations based on aptitude, over the past several decades, a greater emphasis has been placed on measuring achievement for purposes of school accountability. Indeed, one of the paramount goals of NCLB was to create a clearer picture of student achievement by providing detailed test data for individual schools and student subgroups. The hope was that the pursuit of higher test scores would improve instruction and produce real gains in student achievement.
To comply with this mandate, each state adopted its own learning standards, arranged for its own standardized tests, defined its own criteria for proficiency, and used test data to evaluate individual schools and, in some cases, individual teachers. (About half of the states also use standardized exams for “student accountability” in the form of high-school exit exams, as part of a student’s course grade, and for decisions relating to grade promotion or school placement.6) Unfortunately, these assessments generated achievement measures of questionable validity and provided little useful information for teachers. Metrics based on state exams often varied widely from results on the National Assessment of Educational Progress.7

Another rationale for standardized testing is the need for credible indicators of a student’s readiness to succeed, either academically or in the workplace. Although most colleges consider an applicant’s GPA in the admissions process, they must contend with the lack of consistent classroom grading policies. Teachers vary widely in terms of what tasks they grade, what rubrics they use, and how they define the level of rigor associated with a particular grade. The school where I most recently taught adopted a set of policies designed to bring more consistency to classroom grades by using common summative assessments, common task weighting, and other policies intended to minimize variation among teachers of the same course. Even with this proactive approach, it was difficult to ensure that student mastery was measured consistently. Studies have also found that grade inflation has become widespread over the past three decades.8

This lack of consistency and comparability has made it difficult for colleges and businesses to understand what a high-school GPA signifies. As a result, most colleges use SAT or ACT scores, and more-selective schools consider AP or IB exam results. Colleges have also developed their own internal testing policies for course placement, although a recent study found widespread inconsistency in how colleges determine which students require remedial coursework.9

This hodgepodge of testing programs, grading practices, and graduation requirements has made a high-school diploma akin to a devalued currency. Just as monetary inflation and poor fiscal discipline often result in currency devaluation, grade inflation and vague academic rigor have
made a high-school diploma less accepted as a meaningful indicator of performance. As a result, standardized tests are used to fulfill the need for credibility, consistency, and comparability.

Unfortunately, the tests are necessarily limited. Test designers must contend with tradeoffs among the parameters of scope, validity, reliability, and cost. Well-designed tests can measure certain types of proficiency but are less useful in measuring other personal attributes (curiosity, creativity, perseverance, interpersonal skills, and the like) that are essential for successful careers. As the Partnership for Assessment of Readiness for College and Careers (PARCC)—one of the two test consortia that are designing Common Core–aligned tests—acknowledged:

It must be noted that the academic knowledge, skills, and practices defined by the PARCC [College and Career Ready] determinations . . . do not encompass the full range of knowledge, skills, and practices students need for success in postsecondary programs or careers. . . . A comprehensive determination of college and career readiness . . . is beyond the scope of the PARCC.  

Some of these limitations result from the choice of question format and the need to minimize costs. In the United States, standardized tests usually employ a multiple-choice format to increase objectivity and reduce costs. Unfortunately, this format encourages guessing, facilitates cheating, and limits the types of proficiency that can be assessed. With recent improvements in software, some standardized exams have begun to migrate toward constructed-response formats that provide a better measure of a student’s problem-solving skills. Even with these improvements, computer-scored test questions rarely provide the specificity needed to pinpoint the cause of an incorrect response. Therefore, teachers use a variety of supplemental tasks to obtain diagnostic information and a more complete picture of student proficiency.

Another critical shortcoming of standardized assessments, as currently implemented, is their limited ability to provide useful, timely, and actionable information for improving teaching and learning. Well-designed assessments produce feedback that helps teachers and students check for understanding. Classroom teachers routinely rely on formative assessments for this feedback, and PARCC recently announced plans to offer some nonsummative (formative) assessments.
Grant Wiggins, a thought leader in the field of feedback and learning, has identified several attributes of useful feedback; in particular, it must be timely, actionable, tangible, consistent, and transparent. Standardized testing is lacking in several of these dimensions. Other than the overall score, standardized tests provide very little specific feedback to the student or parent, and yield relatively little actionable data for the teacher. Worse still, the results arrive too late in the school year to provide an opportunity to adjust instruction or guide individualized remediation. The annual cycle of mandated exams has become a significant distraction from the learning process rather than a resource that improves it, while test prep has crowded out other learning activities.

Standardized assessments have the potential to play an important role in measuring and improving student achievement. However, the approach currently used by most states does not generate those benefits. We need to reconsider how these assessments can provide more timely and targeted information, consume less time, provide the proper incentives, support personalization, and guide intervention.

**Unique Challenges Associated with Learning Math**

In addition to the limitations posed by standardized assessments, the unique nature of learning math demonstrates the need for a better way to monitor student learning. Acquiring proficiency in math is a long-term, cumulative process that presents special challenges for both the teacher and student. Each phase of learning adds to a foundation upon which the subsequent acquisition of new concepts, skills, and problem-solving strategies relies. This sequential dependency makes math fairly unique compared to other subjects. (For example, what a student learns in US History as a high-school junior is not very dependent on what that student may have learned in World History as a sophomore.)

In 2008, the National Mathematics Advisory Panel recognized the need for a “focused, coherent progression of mathematics learning.” Likewise, the CCSS for Mathematics (CCSSM) states that “what students can learn at any particular grade level depends upon what they have learned before. . . . No set of grade-specific standards can fully reflect the great variety in abilities,
learning rates, and achievement levels of students in any given classroom.”\textsuperscript{15} For many schools, including the ones where I have taught, this heterogeneity is the rule rather than the exception. The aspiration of college and career readiness for all must be reconciled with the reality that teachers confront on a daily basis.

Consider an Algebra 1 class, which is the foundation for subsequent math courses, many technical disciplines (like science and finance), and middle-skill occupations (including nursing, plumbing, welding, and carpentry). Most of the students taking this course are in 8th through 10th grade and have completed a progression of elementary- and middle-school courses where they were taught basic arithmetic and pre-algebra concepts, operational algorithms, math-related vocabulary, and problem-solving strategies. The CCSSM contains more than 200 standards for kindergarten through eighth grade and categorizes them into 11 domains. (See table 1.)

| Table 1. The Common Core State Standards for Mathematics (Number of Standards by Grade Level and Domain) |
|-------------------------------------------------|---------------|----------------|---------------|---------------|---------------|---------------|---------------|---------------|
|                                                   | Kindergarten | First          | Second         | Third          | Fourth         | Fifth          | Sixth          | Seventh        | Eighth         |
| Counting and Cardinality                         | 7            |                |                |               |               |               |               |               |               |
| Operations and Algebraic Thinking                | 5            | 8              | 4              | 9             | 5             | 3             |                |               |               |
| Number and Operations in Base 10                 | 1            | 6              | 9              | 3             | 6             | 7             |                |               |               |
| Numbers and Operations–Fractions                 |              |                |                |               |               |               | 3             | 7             | 7             |
| Measurement and Data                             | 3            | 4              | 10             | 8             | 7             | 5             |                |               |               |
| Geometry                                         | 6            | 3              | 3              | 2             | 3             | 4             | 4             | 6             | 9             |
| Ratios and Proportional Relationships            |              |                |                |               |               |               | 3             | 3             |               |
| The Number System                                |              |                |                |               |               |               | 8             | 3             | 2             |
| Expressions and Equations                        |              |                |                |               |               |               | 9             | 4             | 8             |
| Functions                                        |              |                |                |               |               |               |                |               | 5             |
| Statistics and Probability                      |              |                |                |               |               |               | 5             | 8             | 4             |
| **Total**                                        | **22**       | **21**         | **26**         | **25**        | **28**        | **26**        | **29**        | **24**        | **28**        |

At various points along this path to Algebra 1, a student may fail to comprehend a fundamental concept, key vocabulary term, important operational procedure, or problem-solving strategy. If this proficiency gap is not quickly identified and remedied, that student’s subsequent achievement will be impaired. For immigrant teenagers with limited educational backgrounds, the equivalent of elementary- and middle-school math courses is often compressed into two to three years, and during this time, the student is usually in the early stages of acquiring academic fluency in English. He or she therefore often begins Algebra 1 with a very weak foundation.

Consider these two math questions from the June 2014 New York Regents Algebra 1 (Common Core) Exam:

26. The breakdown of a sample of a chemical compound is represented by the function \( p(t) = 300(0.5)^t \), where \( p(t) \) represents the number of milligrams of the substance and \( t \) represents the time, in years. In the function \( p(t) \), explain what 0.5 and 300 represent.

36. An animal shelter spends $2.35 per day to care for each cat and $5.50 per day to care for each dog. Pat noticed that the shelter spent $89.50 caring for cats and dogs on Wednesday.
   - Write an equation to represent the possible numbers of cats and dogs that could have been at the shelter on Wednesday.
   - Pat said that there might have been 8 cats and 14 dogs at the shelter on Wednesday. Are Pat’s numbers possible? Use your equation to justify your answer.
   - Later, Pat found a record showing that there were a total of 22 cats and dogs at the shelter on Wednesday. How many cats were at the shelter on Wednesday?

To solve these types of problems, a student must be able to:

1. Comprehend the question wording, which is densely packed with essential information (a particular area of difficulty for English-language learners and those with reading disabilities);
2. Determine the relevance of the information provided, the context of the numerical values, and the question(s) being posed;
3. Understand the math terminology and notation;
4. Convert information from one form of representation to another (including verbal, symbolic, graphic, and tabular forms);
5. Use a variety of algebraic and arithmetic methods to model the given facts and compute numeric solutions; and
6. Provide a verbal explanation of one’s reasoning and conclusions.

The integrated nature of mathematical problem solving presents a significant challenge for students with weak procedural fluency, limited vocabulary, and gaps in their conceptual understanding. The widespread use of multiple-choice questions in testing math has masked some of these learning deficits because students often simply guess or use a plug-and-chug strategy to select an answer.

I can personally attest to the difficulty of teaching high-school math when students enter underprepared. My colleagues and I frequently lamented our students’ weak mathematical foundation and its impact on their ability to grasp new topics and develop more-advanced skills. Considerable class time was consumed reawakening prior knowledge and re-teaching skills that were not mastered in previous courses.

Currently, teachers do not have an efficient way to identify specific learning gaps from prior years. A typical high-school teacher is assigned about 120–140 students every year, without a scheme to provide meaningful background information about each student’s math learning history. A better system would offer teachers both a barometer for measuring current-year learning and a mechanism for identifying the most significant gaps in each child’s math learning history with sufficient detail to support intervention.

Another impediment to the instruction-assessment feedback loop is the pacing of the traditional school calendar. In most schools, students are enrolled in a math class for approximately nine months during which they are expected to master the standards associated with that course. Students with learning disabilities may be assigned to a team-taught or self-contained class.
English-language learners may be placed in a general-education class if their math foundation is deemed adequate.

Although standardized-test results for various subgroups are tabulated separately for accountability purposes, all students are essentially put on the same learning treadmill and expected to run at the same pace. The teacher is expected to provide instruction that challenges the more advanced students, remediates the weaker students, and accommodates those with special needs or limited language proficiency. For learning math, this is a recipe for failure.

By attempting to serve everyone at the same time and the same pace, the curriculum serves very few well. The most proficient students gravitate toward advanced courses and progress more rapidly, but for many students in the forgotten middle, the heterogeneity of the class can be a barrier to effective learning, unless the school provides an option for more personalized instruction. And while instructional differentiation is strongly encouraged, assessment differentiation is usually restricted (except for Individualized Education Program accommodations). This disconnect may explain why a recent ACT study found that “despite early detection of academic struggles, students were seldom able to close the gap as they progressed through school.”

Those who do not work in a K–12 school cannot fully appreciate how the modern era of annual high-stakes testing influences the flow of a school year. In a typical high-school math course, the teacher presents a sequence of units (most courses consist of about 10 units, each of which has 6 to 8 lesson blocks). Within each unit, a variety of classroom activities and homework assignments are used to convey key concepts and strengthen students’ procedural fluency.

At the end of each unit, a summative test is administered. Students who are absent for any reason (such as illness, field trips, or truancy) must exert considerable effort to catch up; otherwise, they quickly fall behind and are less prepared for subsequent units. There is very little slack in the schedule; it is a race against the clock to cover all of the standards before testing season begins. Over the past decade, many opportunities for deeper learning have been dropped to focus on topics likely to be tested.
By late April, the test-preparation activities begin. Lessons are used to review some topics from earlier in the year and acquaint students with standardized-test question formats. The school’s information technology (IT) staff begins a herculean effort to install large-scale computer testing facilities in gyms, libraries, and meeting halls. After the AP and IB exams are completed, the state exams begin. School bell schedules are modified, and for weeks, the entire school day is dominated by whatever exams are taking place. Most high-school students take two or three state exams each year and spend two to four hours on each one.

As testing season progresses, students miss many regular classes that overlap with their test sessions. Some students are required to make up or retake a test if they missed the original test date or failed to achieve a passing score. By mid-June, most state exams are completed and the course final exams begin. In total, about 15 to 20 percent of the school year is consumed with preparation for and administration of end-of-course tests.

Although teachers may receive summary score reports within a few days (for tests administered online), very little attention is given to anything except the passing rate. The feedback included in these score reports is too vague to be of value for remediating students who need to retake the test; instead, schools often use “cramming and jamming” sessions in an effort to increase their school’s pass rates. More detailed reports may arrive during the subsequent school year, but even these provide little actionable feedback. Some test questions may be released by the state and used to prepare students for the next testing season.

After all this effort, the end of the year becomes a predictable series of events. The higher-achieving students earn a respectable grade, and some may move up to an honors course. The lowest-achieving students fail and repeat the course the following year. For some students in the middle, many teachers (with the encouragement of administrators who are eager to boost graduation rates) grant a barely passing grade, and the students progress to the next course.

A study of New York Regents Exam scores revealed a suspicious pattern of exam scoring for those students whose scores were very near the lowest-possible passing score. In another recently publicized incident, more than 80 percent of the high-school students in Montgomery
County, Maryland (a highly regarded school system), failed their Algebra 1 final exam, although only a small percentage of students were given a failing grade for the course.19

The current approach to standardized testing has adversely affected student learning in math. It has depersonalized the pacing of instruction, made it more difficult to adapt to the needs of English-language learners and those with learning disabilities, and crowded out time that could be used to provide a deeper understanding of and greater appreciation for the relevancy of mathematical concepts. At the same time, technology is not being fully utilized to provide teachers and administrators with useful and timely information.

A New Way of Doing Things
And yet, it does not have to be this way. If a new approach were designed that could attain the benefits of computer-based standardized testing while minimizing the drawbacks, it would likely rely on a few core principles. Rather than use a single annual standardized test to measure achievement, administering frequent, hybrid exams could help integrate traditional classroom-summative tests with state standardized tests.

Meaningful assessment data would be provided in a timely manner to teachers, administrators, students, and parents. More class time would be spent on instruction, diagnostic assessments, targeted remediation, enrichment, and in-depth investigations. Pacing of both instruction and assessment could become more personalized to ensure that students acquire and maintain a sufficient level of mastery as they progress through the course sequence. The tests would leverage the most recent advances in software to provide better indicators of student proficiency and would allow teachers to spend more time helping students who need individualized attention.

These principles reflect some of the recommendations of the Gordon Commission on the Future of Assessment in Education, whose 2013 report described how assessment policies, strategies, and practices need to adapt to a rapidly changing world.20 James Pellegrino, the commission’s cochair, emphasized this theme in a recent interview:
We need to focus on assessment for learning at the classroom level. We are so hung up on assessment of learning for purposes of accountability or monitoring that we’ve lost track of the investment we need in quality resources . . . to help teachers help students.21

Other scholars have also envisioned an approach in which standardized assessments are disaggregated and synchronized with instruction.22 Implementing such an approach would require six key components.

1. **Test-item authoring:** Teachers could use software to create computer-scorable questions to be included in the hybrid-classroom summative tests. The software would easily integrate textual and graphical components and would support a variety of question formats. Questions with numeric solutions could be customized through embedded algorithms. Ideally, this software would be fully compatible with that used to create standardized questions so that hybrid-test construction could be simplified. (One major challenge is the user interface of the question format, which I discuss in more detail later.) Ultimately, this tool could allow teachers to create computer-scorable formative assessments so that students could become familiar with the question formats, and all assessments would be clearly aligned with relevant standards.

2. **A centralized test-item bank:** Standardized test items aligned with state standards would reside in a central repository. They would be developed by one or more existing test providers, some of whom already offer test-item banks as supplements to their formative assessment services. Each test item would be tagged by standard, topic, cognitive complexity, linguistic level, question format, and other relevant descriptors. A few states have already begun to explore the possibility of using individual test items developed by PARCC or Smarter Balanced Assessment Consortium, while retaining the prerogative of implementing their own standardized exams.

3. **Item review committee:** Each item in the central repository could be vetted by a review committee composed of faculty representatives (such as the American Federation of Teachers, National Education Association, and National Council of Teachers of
Mathematics). This would provide an additional level of quality control and give teachers greater assurance that test items are fair, well written, and aligned with standards. There are several precedents for the role of an item review committee, including the development of the Massachusetts Comprehensive Assessment System. By giving teachers an opportunity to review test items and express concerns, the assessment environment could become less confrontational.

4. **Hybrid-test construction:** Teachers could construct hybrid unit-summative tests that combine their own test questions with items retrieved from the centralized test bank. (Some textbook publishers already offer a similar tool.) Ideally, the test’s format would seamlessly integrate questions from the two sources. Over the course of a typical year, a student could be presented with about 150 items from the central database, compared to about 50 questions used in a typical annual standardized test. States could establish policies to specify what portion of these hybrid tests consist of standardized items, the teachers’ role in previewing questions, and the level of rigor required for each hybrid test.

Such policies would lay the foundation for the software-based algorithms used to construct the individual hybrid tests. Students would be motivated to answer each question on the assessments, since all of their answers might contribute to both their test score and course grade. (This might help alleviate a current concern that students and teachers have different stakes in the standardized-testing process.) Incorporating a larger set of standardized items into the year-long assessment process would improve validity and enable more detailed proficiency metrics.

5. **Automated test scoring and cut scores:** Computer software would score each student’s exam and provide teachers with sufficient details to support classroom grading. At the same time, the responses to the standardized-test questions would be retained in a secure database maintained by the state. States would establish policies to set the level of proficiency needed for diploma eligibility, including an approach for aggregating test results over multiple years. Student progress would be monitored on a more frequent basis instead of using annual high-stakes test sessions to determine student status.
6. **Test-data analysis:** As the proficiency data is accumulated, analytics software could provide authorized users with customized reports that contain actionable information. Classroom teachers would have dashboards to identify specific areas of student weakness and could receive a detailed, customized report at the beginning of each school year that would provide a meaningful profile of each student’s past proficiency. This report might also be used by school administrators to support ability grouping, personalized pacing, or targeted intervention strategies. It would facilitate the longitudinal study of proficiency growth and enable comparisons of classroom grades with standardized-testing metrics. A well-designed analytics-based system could also improve the productivity of faculty and administrators. State and local authorities would control data access to ensure that student privacy concerns are addressed and federal regulations followed.

This assessment infrastructure offers several advantages over the status quo while preserving the authority of, and current relationships among, federal, state, and local authorities. It is intentionally designed to be flexible with respect to standards and curricula, and as long as the centralized question bank contains appropriately aligned items, each state could use its own standards. The system can be adapted to a variety of curricula, course-pacing guides, and classroom assessment policies because each test item is individually tagged. Although the system as described here envisions a hybrid test for each course unit, some schools may prefer to give the hybrid tests only on a quarterly basis. And, for example, some states may decide to use the hybrid approach for middle school but not high school.

Although I believe that more uniform standards and more consistent cut scores would be beneficial, these should not be prerequisites for adopting hybrid exams and longitudinal proficiency tracking. The concepts and laws of Algebra are the same for those living in Wisconsin and Wyoming, but variations in overall state standards would not preclude comparability for individual standards (such as CCSS clusters) that are shared among multiple states. As long as the federal government agreed that modularized, hybrid assessments can generate the data needed to satisfy federal accountability mandates, states and local authorities could continue to exercise their powers to set policies and oversee their implementation. If well
implemented, standardized tests could become woven into the fabric of a school rather than
dropped from the sky each spring.

**Challenges.** Although this assessment environment would enable more timely and useful
feedback, state agencies would still need to decide what level and mix of proficiency would be
required for students to be eligible for a diploma. This flexible structure does present a variety of
implementation challenges.

*Legal and Contractual Issues.* Almost all states currently contract out many tasks related to
creating and administering standardized exams. These contractors are developing different test
items, different test-delivery platforms, and somewhat different test characteristics. The hybrid-
test concept described here could be implemented independently by the two Common Core
consortia (PARCC and Smarter Balanced) or by testing companies in response to guidelines
issued by state authorities. Alternatively, these organizations might decide to collaborate on the
related technology development or test content.

*Technology Development.* The successful adoption of any technology-based innovation will
depend on software that is truly user friendly and designed to meet the needs of teachers and
administrators. Several companies have already developed test-item authoring tools, test
generators, online test-administration modules, test-scoring algorithms, and data-analytics
capabilities. However, these resources currently exist in separate, proprietary product offerings,
and the tools used to create and manage classroom tests are not interoperable with those used for
states’ standardized exams. Some of the intellectual property owned by various private
enterprises may need to be licensed or shared for this integrated concept to meet its full potential.

*Usability.* The wider use of the constructed-response format presents several significant user-
interface challenges. Some of the new formats are fairly straightforward, but others pose
significant man-machine interface challenges that involve using a computer mouse or virtual
keyboard and toolbar to write math equations, graph functions, perform geometric constructions,
or draw diagrams on a Cartesian grid. To fully understand this issue, one must experience it
firsthand.
The blog at numberwarrior.wordpress.com has several postings that clearly illustrate some of the issues in the online practice tests released by PARCC. Unless students have time to become familiar with these interfaces and adept at using the online tools, their test results may not accurately reflect their abilities. This is an even bigger risk when the entire year’s learning is measured in a single testing session. The hybrid-testing concept would provide more opportunities for students to become familiar with these interfaces, and the stakes for any single test session would be reduced.

**IT Infrastructure.** The approach outlined earlier envisions the use of computer infrastructure to deliver the tests and maintain a database of results. Many schools currently lack the needed capacity, expertise, network connectivity, and operational support resources to execute this concept immediately. However, over the next few years, as the next generation of online testing is rolled out, more schools will acquire the capabilities to implement it.

**Test Security and Privacy Issues.** As with existing standardized-testing programs, the need to secure test content and protect the privacy of student data is paramount. Any solution will need to comply with federal regulations and guidelines regarding the use and dissemination of student data, which should be controlled by state authorities to avoid problems that impaired previous attempts to outsource this function. School leaders should develop strategies for clearly explaining to parents the benefits of using information to improve learning and proactively address concerns about potential misuse of student data.

**Test Validity and Reliability.** The hybrid-testing concept raises significant challenges related to test validity and integrity. Testing sessions would be more frequent and more decentralized (for example, administered in individual classrooms). Policies and procedures would need to be developed to ensure that the integrity of the assessment process is maintained and that testing practices provide valid and reliable measures of student proficiency. One advantage of the hybrid approach is that data is collected from a larger set of test questions, which could provide more valid proficiency measures than annual tests that cover an entire year.
Policy Implications
Although the 2001 passage of NCLB was a bipartisan effort, support weakened over the next decade as a variety of program flaws became more apparent. More recently, the pushback against the CCSS rollout illustrates how sweeping changes to standards and assessment can become extremely controversial. The hybrid-testing approach outlined here attempts to address some of these concerns and explain how a more robust assessment infrastructure could be used to support personalized curricula and foster more thoughtful diploma-eligibility policies.

A few states—including New Hampshire, Maine, and Oregon—have announced their intent to migrate toward a competency-based curriculum in which students will learn at a more personalized pace and be required to demonstrate proficiency before progressing to more advanced units. This approach often incorporates a blend of online tutorials and assessments, in-depth investigations, and longitudinal performance tracking. However, a report from CompetencyWorks identified a major regulatory issue:

An outdated federal accountability system presents significant barriers to widespread implementation of competency education. . . . the system’s focus on after-the-fact summative tests, time-based elements, annual rankings, and narrow indicators of success have created disincentives for educators who are interested in student-centered learning and accountability policies focused on individual student growth, pace, and achievement. Without federal action, competency-based educators are compelled to maintain parallel accountability systems; one maintained by federal law and one that aligns with their vision of success for every student.23

The hybrid-testing concept would alleviate much of this concern and avoid the duplicative expenditure of time and money on separate assessment requirements. Students would no longer be held hostage to the annual cycle of state testing, because those assessments would be modularized and incorporated into the course units.

The debate over the CCSS has highlighted some even-more-fundamental questions regarding learning standards. Are “college-readiness” and “career-readiness” equivalent descriptors? More
broadly stated, we should be asking what types of math proficiency are essential for all students. The OECD Survey of Workplace Skills, Technology, and Management Practices found that only a small fraction of the American workforce uses advanced high-school or college-level math (see figure 1).

**Figure 1. What Percentage of Americans Actually Use Math at Work?**

![Bar chart showing the percentage of Americans using math at work](chart)


Similarly, the National Center on Education and the Economy studied the specific math skills used in many community-college programs and found that relatively few students needed math knowledge beyond a course they labeled “Algebra 1.25” (Algebra 1 plus a few topics in statistics, discrete math, and geometry).\(^{25}\) (Figure 2 shows the specific math topics used in business math courses.)
Figure 2. Average Percentage of Text Chapters and Exam Items Containing CCSSM Domains for Business Math Courses


Although the mixture of math topics varied among the different programs, the report concluded:

> In sum, a substantial part of the high school mathematics we teach is mathematics that most students do not need, some of what is needed . . . is not taught . . . and the mathematics that is most needed . . . is actually elementary and middle school mathematics that is not learned well enough by many to enable them to succeed. . . . It makes no sense to rush through the middle school mathematics curriculum in order to get to advanced algebra as rapidly as possible.\textsuperscript{26}

The Southern Regional Education Board examined this issue from the perspective of students who pursued career and technical courses in high school. It recommended that “students pursuing a concentration in specialized career and technical (CT) studies could be offered an
option to pass state-approved, nationally recognized employer certification examinations in lieu of passing additional academic examinations.”

Perhaps of gravest concern, a June 2014 study by the American Education Research Association suggests that increasing math course graduation requirements could have adverse effects on high-school dropout rates and college enrollment, particularly for some at-risk student subgroups. The researchers concluded that “policymakers must anticipate unintended consequences from more demanding content and more rigorous requirements.”

Today, too many students are progressing through high school with a very weak understanding of basic math concepts while being required to learn more-advanced topics related to college-level science, technology, engineering, and math programs.

We should be placing greater emphasis on the deeper mastery of essential, widely applicable mathematical concepts including pre-algebra, basic algebra, basic statistics, and probability. Once students have demonstrated true mastery of these core topics, they could either progress to more-advanced math courses or take career and technical courses that apply these topics and perhaps develop more marketable skills. The hybrid-assessment infrastructure would allow schools to measure and monitor student proficiency with greater precision, frequency, and scope. It would enable mastery subscores to be developed for the most important math topics rather than simply using a one-shot annual test to assign a single numeric value to each student’s overall achievement.

Initial results from the first wave of CCSS-aligned tests foreshadow the dilemma states will face when they attempt to reconcile the more rigorous assessments with the desire to maintain and improve graduation rates. For example, New York adjusted the 2014 Algebra 1 exam cut score (30 out of 86 points) so that pass rates were in line with historical norms, even though that level of mastery did not meet CCSS readiness standards. A New America Foundation report also predicted that several states may begin to adopt two cut scores: one for high-school graduation purposes and another to indicate college readiness. We need to give more careful thought to the way we assess student mastery so that the resulting proficiency indicators are relevant to the full spectrum of postsecondary occupations.
Conclusion

Our nation’s K–12 educational system is plagued with a number of challenges, and there is no silver bullet that will remedy all of its shortcomings. Educational achievement ultimately depends on a motivated student, supportive home environment, effective teacher, well-structured curriculum, and competent administrators. Assessments can play a vital role in providing the type of feedback that helps students develop their skills, helps teachers adapt their instructional strategies, and assists administrators in efficiently allocating resources.

Advances in science and medicine depend on calibrated instruments that provide reliable and accurate measurements. Successful businesses frequently monitor their performance and take corrective action in response to that feedback. It is time schools and policymakers learn from these other sectors how information, when properly gathered and interpreted, can yield better outcomes. With a more robust assessment infrastructure in the American K–12 education system, we may begin the long, gradual path toward a better-educated workforce and a more informed citizenry.

Steven Shapiro (steve.shapiro31@gmail.com) recently retired from Fairfax County Public Schools (Virginia), where he was a high-school math teacher. Before his teaching career, he served as a management consultant and corporate executive for several high-technology companies.

Notes


4. According to Falls Church High School’s enrollment profile, 58 percent of students receive free or reduced lunch, 27 percent have limited English proficiency, and 18 percent qualify for special-education services. See Fairfax
Assessment

