



Starr



Spellings

Examining High-Stakes Testing

Education Next talks with JOSHUA P. STARR and MARGARET SPELLINGS

More than 40 states plan to assess student performance with new tests tied to the Common Core State Standards. In summer 2013, results from Common Core–aligned tests in New York showed a steep decline in outcomes. Common Core advocates hailed the scores as an honest accounting of school and student performance, while others worried that they reflected problems with the tests, inadequate support for educators, or a lack of alignment between what schools are teaching and what’s being tested. In this forum, Joshua Starr, superintendent of schools in high-performing Montgomery County, Maryland, makes the case for a three-year hiatus from high-stakes accountability testing while new standards and tests are implemented. Accountability proponent Margaret Spellings, U.S. secretary of education from 2005 to 2009 and now president of the George W. Bush Presidential Center, defends the testing regime as a critical source of information, for educators as well as the public, and argues for holding the line.

A Testing Moratorium Is Necessary

by JOSHUA P. STARR

Great instruction needs great assessments.

Meaningful assessment data reveal what students know and are able to do, and provide teachers with the information they need to track student progress and to identify and support students who are struggling. Assessment data give central-office administrators and school boards the crucial information they need to allocate and evaluate resources effectively and to set policies.

So why do we need a three-year moratorium from accountability systems based on state tests? At this crucial time in American public education, when we are correctly focusing our attention on the rigorous Common Core State

(continued on page 72)

Assessments Are Vital for Healthy Schools

by MARGARET SPELLINGS

Putting a moratorium on testing is akin to shooting the messenger.

Standards, tests, and accountability policies are merely tools. They don’t make learning happen. Tests themselves don’t narrow the curriculum; they also can’t close achievement gaps. How educators use these tools is what is critical. Superintendent Starr argues that testing and accountability are important for “developing indicators that can inform an organization’s actions.” But his emphasis on assessment and accountability as tools for managing education bureaucracy is only part of the story—and cold comfort for families who want and need the information to access the

(continued on page 73)

Standards (CCSS), we must organize for success in the future and not remain fixated on the past. Critics of a moratorium and defenders of the status quo say my approach jettisons accountability for schools. Nothing could be further from the truth. We must build systems of accountability and support that use the right assessments to measure the right things. Accountability and support can often be seen as competing demands, but comprehensive assessment data actually serve both functions.

I started my administrative career in 1998 as director of accountability for a small district, and later served as the director of school performance and accountability for New York City public schools. In Montgomery County, Maryland, where I am now superintendent, we are on our way to building an accountability system that uses student data in meaningful ways to measure progress and improve instruction at important checkpoints along a child's educational journey. Having designed accountability systems for different types of districts, I recognize the importance of developing indicators that can inform an organization's actions.

Current assessments, unfortunately, do not measure what our students need to know and be able to do in the 21st century. Basing decisions on these outdated state test-score data may lead to structural changes that seek to address the wrong problems.

The Legacy of NCLB

Public education is extremely complex. Multiple entities govern, drive, and constrain the work of educators, from federal and state laws and regulations to local political structures, funding authorities, and interest groups. Effective district and school leaders must mitigate the clamor of competing interests and demands, and focus on organizing teaching and learning systems around what really matters. State and federal accountability systems should provide and enhance that focus, not distract from it.

No Child Left Behind (NCLB) provided this focus for public schools over the past decade, whether we like to admit

it or not. Regardless of lofty mission statements that spoke of meeting the needs of the whole child, cultivating artistic curiosity, and having high academic standards, every school in America has had one primary mission since 2001: to meet Adequate Yearly Progress (AYP). Curriculum was narrowed and aligned to tests; professional development centered on data-based decisionmaking; supports and funding were put in place to improve test scores; data systems provided information for school and district leaders to make decisions about policy and allocating resources; and political entities had evidence to celebrate or complain about the investment of tax revenue.

NCLB did a good job of making data an important part of the school improvement process and of exposing the persistent achievement gaps at even the most high-achieving schools. But these data are based on assessments that are very limited in what they measure and don't reflect the skills and knowledge our students need to be successful. Moreover, the goal of 100 percent proficiency may properly reflect the desire to ensure all children achieve, but it's not realistic. It's akin to saying that a person is only physically fit if she can run a marathon. NCLB's goal was "adequacy." Now we need to develop measures that will tell us whether our children will thrive in a 21st-century economy and world.

NCLB is dying a slow death. We have entered a new era of American public-education reform, brought on by Race to the Top and the Common Core State Standards. In Montgomery County, we are creating new systems that holistically measure whether a school is supporting a student's academic success, creative problem-solving abilities, and social and emotional well-being. This requires a different approach to teaching and learning, and to supporting and holding schools accountable.

New Tests for a New Standard

The major change happening in public education today is the democratization of information. As we all have greater access to information than ever before, educators must enable students to create knowledge and wisdom from the information surrounding them. We have to rethink what we're asking teachers, support professionals, and leaders to do, and then build the organizational structures to support them.



Continued
page 74



The goal of 100 percent proficiency may properly reflect the desire to ensure all children achieve, but it's not realistic.

best education for their children now.

The goal of No Child Left Behind (NCLB) has been to put the focus of education policy squarely on students. Most importantly, then, test results provide parents and teachers with vital information about student learning, and accountability policies challenge



districts and schools to meet individual student needs with effective teachers, strong curricula, choices for families and students, and break-the-mold interventions for failing schools. For over a decade now, test-based accountability has acted as a sort of insurance policy to make sure disadvantaged and struggling students are not ignored. I take Superintendent Starr at his word that his proposed moratorium on testing and accountability would be temporary. But I am skeptical. After more than a decade of resistance to NCLB by the education establishment, I find something disingenuous about the argument that schools ought not to be held accountable to the standards states themselves set for grade-level student achievement. Helping all students read and cipher on grade level is a modest goal for our children and grandchildren. As long as a significant portion of students aren't reaching these so-called "outdated" state standards, we must continue to assess the skills and hold schools accountable for the results. States' efforts to ensure college and career readiness for all depend on it.

Testing Is Critically Important

Because of assessments, we can track the academic progress of American students. Recent results on our Nation's Report Card (the National Assessment of Educational Progress, or NAEP), for example, tell us that during the NCLB era, student achievement in reading and math improved for African American, Hispanic, and white students alike, and achievement gaps among these groups narrowed. As Paul Peterson recently pointed out in the *Wall Street Journal* (August 7, 2013), between 1999 and 2008, on the NAEP, white nine-year-olds gained 11 points in math, African American students gained 13 points, and Hispanic student performance improved by 21 points. In reading, white nine-year-olds gained 7 points, black performance jumped by 18 points, and Hispanic scores climbed 14 points. (Importantly, Peterson also notes that gains have diminished since the Obama administration began to dismantle NCLB.)

Tests also identify where we are falling short. For example, despite significant progress, NAEP scores reveal that just 34 percent of our nation's 8th graders are proficient in reading

No one has ever demonstrated that mastering grade-level reading and math skills hurts students' ability to acquire higher-order thinking skills.

and 43 percent are proficient in math. The achievement of students in American high schools, where state testing is minimal (and accountability weakest), hasn't budged in four decades.

Test-based accountability policies have demonstrated unequivocally that what gets measured matters. A recent report by Common Core, Inc., its title intended to demonstrate that students are "Learning Less" because of assessments, included some interesting findings: ninety percent of teachers say that when a subject is included in a state's system of testing, it is taken more seriously. Eighty percent of teachers say that their schools have been offering more "extra help for students struggling in math and language arts" in recent years. This is good news. This is student assessment used to inform classroom practice, which is what it's meant to do. The truly important questions we face in education reform aren't about whether we should test students but rather about how schools will respond to what tests tell us about student needs, and what districts and schools will do differently to ensure that all students learn.

The Testing Critics

Resistance to assessment, accountability, and transparency remains fierce, and not at all temporary.

Critics attack testing from all possible angles, and frankly, the arguments are not particularly coherent. For example, Secretary of Education Arne Duncan claims that states have "dummied down" their standards yet at the same time, his department is giving states waivers to provide "relief" from the unrealistically ambitious expectations of NCLB.

Some argue that the real problem with annual state tests of grade-level reading and math skills is that they force teachers to narrow their focus, distracting teachers from other subjects and the more sophisticated academic skills they would otherwise engender in students.

But no one has ever demonstrated that mastering grade-level reading and math skills hurts students' ability to acquire higher-order thinking skills. Nor has anyone shown that state standards in reading and math endanger students' social and emotional



Continued
page 75

This is a complex conversation. It will require an enormous amount of time and energy to engage adults in new learning about what students should know and be able to do when they graduate in 2025. Along the way, we must tackle some very important questions. What should the classroom look like? What materials and technologies are available to support instruction? What training should teachers undergo? What are the roles of the instructional leader and the central office? How are the community and families to be engaged? What are the right funding mechanisms? What policies and practices should be in place? These are questions that must be addressed at the school, central office, and board and community level, and the answers may very well be different among districts.



In organizing these new systems, we cannot have two areas of focus. State standardized tests—the foundation of NCLB—are not aligned to the CCSS, yet these tests are still being given and are now tied to revised accountability systems under the NCLB waiver program. For example, in math, state tests measure computation skills only, while under the CCSS, students need to show their reasoning ability in addition to computation skills. While we are in this transition, we are telling our teachers, leaders, and authorizing, funding, and governing agencies that we've got to prepare for a whole new standard that will make us more competitive internationally. So where should we focus our attention? How do I explain to my principals and teachers that the current state tests are meaningless because they assess an old standard, even though they are administered, used for accountability purposes, and reported in the media? How do you ask people to work harder than ever to learn new methods of teaching and learning, design new data systems, and invest in new technologies when their evaluations are being tied to the old measure under Race to the Top? How do we determine an appropriate rate of change, one of the most confounding leadership decisions, one that will enable us to switch, almost overnight, to new testing and accountability systems, while also giving people a chance to learn, grow, and adapt with these new systems? Teacher evaluation systems like the professional growth system in Montgomery County are excellent examples of assessing teachers' skills and competencies without an overreliance on an annual state-administered test.

Managing the Transition

As a school superintendent, I have to balance the need for consistent standards for outcomes and processes with my belief in school-level innovation and creativity. I also have to help our organization transition from a bureaucracy that looks much like it did 25 or 50 years ago to one that performs with speed and flexibility. I have to communicate our progress and our needs to the community, the board of

State standardized tests —the foundation of NCLB —are not aligned to the CCSS, yet these tests are still being given and are now tied to revised accountability systems under the NCLB waiver program.

education, and the local elected bodies that provide funding. I need a starting point for determining how to allocate resources, invest additional time and energy, provide supports, and ensure accountability. It is essential, then, that I have a handful of clear indicators that provide starting points for further analysis.

When I go to the doctor I have my vital signs taken first. I believe we need similar vital signs for public education. In Montgomery County, we are organizing our efforts around five milestones:

- 1) reading on grade level by 3rd grade
- 2) completing 5th grade with the necessary math, literacy, and social-emotional skills to be successful in middle school
- 3) completing 8th grade with the necessary math, literacy, and social-emotional skills to be successful in high school
- 4) having a successful 9th-grade year, as measured by grade-point average, well-being, and eligibility to participate in extracurricular activities
- 5) graduating high school ready for college and career, as measured by such existing indicators as performance on Advanced Placement and International Baccalaureate exams and SAT scores.

These are not the only indicators of an excellent, well-rounded education, but each is an important milestone in a child's education and is a starting point for our team to focus attention. Data for each area need to be analyzed by school and by demographic, socioeconomic, and programmatic



Continued
page 76

well-being. While the narrowing curriculum rallying cry is popular in opinion surveys, assessments such as NAEP reveal no signs of declining achievement in science or history or any other supposedly “squeezed out” subject.

Annual state assessments in Maryland take six hours, the equivalent of just one school day. Yet testing critics would have us believe that the creativity of teachers is completely shackled for the other roughly 179 days of the school year. This argument remains popular, even as teachers report their preparation and ability to teach critical thinking and complex problem solving to be limited, and even while so many schools are achieving not-so-stellar results on what are dubbed as less-than-sophisticated current state tests.

Many critics argue that annual state tests in reading and mathematics are inappropriate because not everything students need to learn can be measured by standardized tests, downplaying what even our so-called “crude” tests reveal about serious gaps in the important skills students need. And it is still not uncommon to hear educators insist that assessing students in reading and math is unfair, especially to students likely not to perform well. You see, the schools are fine; it’s the students of color, students in poverty, special education students, and English language learners who are the problem.

The collection of objections is endless. But all of them evade a simple explanation for why education standards with regular assessments of student progress, transparency for results, consequences for school failure, and choices for families have always been under fire. They demand public accountability for education systems across the nation, and many, many public-school systems and educators in the United States simply reject the concept out of hand.

by race and income. Most such student-achievement targets were established for “reporting purposes only” and are no longer used for any meaningful school accountability purposes. Most states now combine student subgroups, previously identified by race, ethnicity, economic disadvantage, special education, and English language learner status, into opaque “super-subgroups” that are very purposefully less transparent. We are turning back the clock to the days when expecting less from the kids who need our public schools the most was acceptable practice. It is the “soft bigotry of low expectations” President Bush so rightly decried.

The Legacy of NCLB

No Child Left Behind and its state testing mandates have always been maligned by protectors of the education status quo. But our memories are too short. Before NCLB, the education establishment thought it fine, even appropriate, to set different academic expectations for kids based on their ethnicity, zip code, or parents’ income. Overwhelmingly, poor and minority students were denied meaningful educational opportunities because of the abysmal quality of schools they attended. Parents had scant information to compare schools. Taxpayers got little more than an ever-increasing invoice for our schools.

Under NCLB, for the first time, schools were required to measure improvement in student achievement across all groups of students, and each state, district, and school was required to lay the results out on the table for parents and the public to see. Parents now know whether their children are meeting state standards in reading and math and which students are being educated, by whom, and in what schools. Taxpayers now know more about where their dollars are

being invested and what the results are. We have sophisticated data that can be used to improve learning in classrooms in real time. We can do a better job evaluating teachers, informed at least in part by the performance of their students. We can tell how students are performing against a standard, and compare them to students in other schools, districts, and states. We can redirect our

resources to where there is the greatest student need.

We’ve made significant improvements in student achievement but we are far from the finish line. State test data reveal significant achievement gaps yet to be addressed, even in high-spending, high-achieving Montgomery County, Maryland,



Kicking the can down the road on assessment and accountability will neither help close achievement gaps nor prepare students for the Common Core.

If one wants to understand the true interests of the education establishment when it comes to pausing test-based accountability, one only need take a close look at the NCLB waivers given by Secretary Duncan to about 40 states to date. The waivers have allowed states to set race- and income-based goals, that is, to lower expectations for student achievement

Continued
page 76

subgroup. Other data will also be needed to determine whether a school is on the right path.

By focusing on indicators at different stages, rather than every year, we are establishing a developmental approach to accountability. School improvement takes time, and students can blossom over the course of a few years. Schools need to be given the opportunity to grow and develop. But current state tests measure annual performance. School improvement could more effectively focus on the needs of individual students if the message to schools was that their accountability for student achievement was tied to the time they had to improve outcomes.

What happens between these milestones must be the focus of efforts to support a school, while the milestones become the focus of accountability. For example, if one subgroup of students isn't successfully completing 9th grade at the same rates as others, then the school needs to drill down on what's happening with that subgroup. If the school then needs additional resources and support from the district—or is unable to improve outcomes—my team and I would step in. If, over time, a school, with our help, has been unable to improve those outcomes, we would have to employ accountability mechanisms.

Building this type of system takes time, resources, and commitment. Montgomery County Public Schools started working on aligning curriculum to the CCSS four years ago. Even with all of the infrastructure and support in MCPS, it will still take more than two years to fully implement our new accountability system and even longer before all elements of it are effectively used in every school.

A moratorium from state standardized tests tied to NCLB is necessary to allow school districts the opportunity to organize their systems to what we're being asked to do now and in the future: prepare adults to engage students in much deeper learning so they will be equipped not just with the academic skills but with the problem-solving skills necessary to be globally competitive. We have to organize our systems to achieve this goal, which is incredibly difficult when we're still being measured by an outdated model. In the interim, we can continue to measure ourselves by standard indicators of college and career readiness, such as SAT, Advanced Placement, and ACT tests and graduation rates. We could also use a nationally accepted criterion-based reading test to determine our current status, but not for high-stakes accountability purposes.

Once the CCSS is fully implemented and the new assessments aligned to these standards have been completed, we can begin to construct a meaningful accountability system that truly supports teaching and learning. ■

considered by many to be one of the best school systems in the nation. In Montgomery County, 59 percent of white elementary-school students score at what the state defines as the "advanced" level on the Maryland State Assessment in reading, while only 26 percent of African American students can boast the same. On the state math test, 52 percent of white elementary-school students compared to 18 percent of African American students score at the top performance level.

These results don't warrant any kind of hiatus from state testing. Kicking the can down the road on assessment and accountability, in Montgomery County and in school systems across the nation, will neither help close achievement gaps nor prepare students for the Common Core.

The current debate about student testing is misguided. Tests are measurement tools. When I step on my bathroom scale and am not happy with what it records as my weight, it isn't the scale's problem. I can boycott stepping on the scale, or I can decide to examine my lifestyle, determine whether I am exercising too little or eating too much, and come up with a game plan to reach an improvement goal.

Our annual state tests amount to stepping on a scale. If our test results are not what they should be, we need to ask: How are education systems radically reconsidering the way they use their resources to improve outcomes? How are they changing the ways they prepare teachers, pay teachers, organize the school day, use technology for learning, and get our neediest students access to our best and brightest teachers?

Common Core Is No Panacea

More than 12 years have passed since I worked with President George W. Bush, Senator Edward Kennedy, and other congressional leaders to pass No Child Left Behind. Despite the vastly improved information we now have, and that we could use more effectively to improve student outcomes, too many educators remain engaged in wearied debates about whether assessment is an important tool for measuring student learning.

This doesn't bode well for implementation of the Common Core. While all variety of education pundits, reformers, and policymakers discuss the merits of upgrading our public education system to the Common Core and college and career readiness for all, the real battle on the ground is whether educators believe schools are capable of or should be expected to help students meet even basic academic standards.

The very same critics who claim teachers are prevented from teaching higher-order skills because of current testing and accountability policies also argue that we need to put

a stop to testing and accountability because teachers aren't prepared for the higher demands of the Common Core.

I support the Common Core standards. I believe our nation's schools need a challenging and common set of academic expectations that is consistent with the demands of the knowledge economy and global competition. But let's not kid ourselves. It is right to be concerned about whether enough of our nation's teachers are ready for the Common Core. And the Common Core is pie in the sky unless students meet basic grade-level expectations in reading and math, a goal we have fallen woefully short of meeting to date.

This debate just brings out the skeptic in me. I am afraid that we aren't serious. We aren't serious in believing that all kids can learn. We aren't serious about ensuring that poor and minority kids get the education they deserve.

Is learning more than a test score? Of course. Are reading and math all we care about when it comes to student achievement? No. Is there always the promise of better tests and better-prepared teachers down the road? Sure. But does any of this suggest we should we have a moratorium on testing or "hit pause" on school-based accountability? No way. ■



CARTOON / BERNARD SCHOENBAUM / THE NEW YORKER COLLECTION

"I don't have an answer, but you've sure given me a lot to think about."

Program on Education Policy & Governance

Post Doctoral Opportunities:

PEPG offers a one-year resident fellowship at the postdoctoral level starting in Fall 2014.

The fellow will work with PEPG staff at their Harvard University offices and engage in independent research related to the program's focus as well as collaborate with the Program's ongoing research and activities.

For more information visit:

www.hks.harvard.edu/pepg/postdoc.htm

VISIT US ON THE WEB

www.hks.harvard.edu/pepg

New Research: www.hks.harvard.edu/pepg/research.htm

Program Events: www.hks.harvard.edu/pepg/events.htm

