

THE EDUCATION EQUALITY PROJECT POSITION PAPER SERIES ON ACCOUNTABILITY

The Education Equality Project (EEP), like almost every other organization committed to education reform, is “for” accountability in our nation’s schools. Yet favoring school accountability today is often more akin to favoring world peace or justice for all—no one publicly purports to oppose such an amorphous, global goal. In practice, however, the particulars of how students, teachers, schools, districts, and superintendents should be held accountable for raising student achievement are deeply divisive. Precisely because of the rhetorical haze that conceals these underlying policy disagreements, the EEP wants to delineate its position on accountability, which it believes is an essential tool for helping to close the achievement gap.

Despite the common-sense appeal of school accountability, the notion that students, teachers, districts, and superintendents should bear responsibility for raising student performance is a “recent invention” in the words of education historian Diane Ravitch. Prior to the standards movement of the mid-1990s, and the subsequent 2002 passage of the No Child Left Behind (NCLB) law, accountability was an alien concept in most classrooms. While schools had tested students for decades, principals and teachers shouldered little or no responsibility to demonstrate that they were actually bolstering student learning. Schools instead administered IQ testing to assign children to different tracks and to assess what students were capable of learning—rather than to measure academic growth.

While schools and districts collected reams of data in earlier decades, they failed to gather the information needed to create a system of accountability for monitoring student achievement. Educators either collected data to monitor inputs (e.g., spending per student, class size, teacher credentials) or they gathered data for compliance purposes, to fulfill reporting requirements for district, state, and federal programs. Meanwhile, student outcomes were all but ignored. In 1993, Al Shanker, the legendary leader of the American Federation of Teachers, observed pointedly that “unless there is accountability, we will never get the right system. As long as there are no consequences if kids or adults don’t perform, as long as the discussion is not about education and student outcomes, then we’re playing a game as to who has the power [in schools] . . . Unless you start with a very heavy emphasis on accountability—not end with it—you’ll never get a system with all the other pieces falling into place.”

Today, the idea that teachers, principals, and schools should be held accountable for educational outcomes—and that accountability should be a starting point for reform—seems unexceptional. Yet the fact that Al Shanker had to defend the need for a “very heavy emphasis on accountability” is testament to an allergic aversion to accountability measures that

continues to infect much of the education establishment, undermining efforts to narrow the achievement gap. At its root, the ongoing, thinly concealed opposition to accountability stems from the belief that teachers, principals, and school administrators cannot *really* be held accountable for substantially boosting student performance among low-income minority students.

Skeptics of accountability contend that low-income students cannot be expected to make large gains in learning because they bring too many burdens of poverty to the classroom. Teachers, the skeptics argue, cannot be held accountable for raising the performance of students who lag behind when they start kindergarten because their mothers failed to read regularly to them, nor can teachers be responsible for failing to raise the performance of students who had to suffer the upheaval of moving frequently because they were poor. Principals cannot run high-performing schools when students suffer disproportionately from asthma, poor diets, untreated dental and vision problems, and other assorted ills.

Thus, new or expanded anti-poverty programs that boost school readiness, reduce class size, provide rental vouchers to help the poor escape housing projects, and build full-service health clinics in high-poverty schools are the real antidotes to the achievement gap—or so the argument goes. As Diane Ravitch presciently foresaw in 2002 when NCLB was enacted, “policymakers’ pressure for accountability has not run into a brick wall of resistance but a bowl of Jell-O instead, where demands for accountability are eventually but inevitably transformed into demands for more resources. Educators want to improve student performance, but to do so they must have higher salaries, smaller class sizes, more training, and so on.”

The signatories of the Education Equality Project are under no illusions—poverty and its attendant burdens are important impediments to learning, and anti-poverty programs should be pursued to reduce economic hardship among low-income families. But the EEP does not subscribe to the belief that demography is destiny in the classroom, or to the environmental determinism of some educators, who seek to effectively be exempted from being accountable for boosting the academic performance of disadvantaged students.

This anti-accountability ethos has hurt disadvantaged students in at least two ways. First, parents of low-income students have long struggled to find information about better public school alternatives, even though their children currently attend some of the nation’s worst schools. As Andrew Rotherham of Education Sector has pointed out, “it’s ridiculous that today a parent can find more information about choosing a new washing machine or automobile than about choosing a school.” Rev. Al Sharpton, co-founder of the Education Equality Project, perhaps summed up the impact of the accountability shortfall best following a White House meeting on education reform with President Obama in May 2009. Asked afterwards what could realistically change in the next four years in the nation’s schools, Rev. Sharpton responded: “We can begin marking how we close the [achievement] gap—using data, not using excuses.”

The hostility of educators to accountability for student achievement has also led to a cramped understanding of accountability and an accompanying backlash. The first wave of accountability regulation during the standards movement of the 1990s and the NCLB-era has been heavily tilted to collecting snapshot data of student learning, rather than generating more telling longitudinal data. The limitations of snapshot data—combined with NCLB’s emphasis on sanctioning failing schools—helped transform the public perception of accountability in some quarters from a beneficial tool for students and teachers to a soulless form of test prep and fill-in-the-bubble assessments.

Snapshot statistics can provide useful performance data to parents, teachers, and principals about a group of students at a given point in time. Snapshot numbers help to answer questions like, how many eighth graders tested proficient in math and reading at the neighborhood middle school in 2009, or how many seniors graduated from the high school down the street this year? By contrast, longitudinal data allows parents, teachers, and principals to answer questions like, what schools and teachers produce the strongest academic growth for their students when compared to schools with similar student populations? What preschool and early childhood programs have the biggest, or the most lasting, impact on subsequent academic achievement among low-income students? How should instruction be tailored to address the needs of struggling students—and what courses should students take to ensure that they are college or career-ready when they graduate? “Longitudinal statistics,” the Data Quality Campaign sums up, “are better suited for supporting the systemic, long-term effort needed to get large percentages of disadvantaged students ready for college and skilled careers.”

TOWARD A NEW “CULTURE OF ACCOUNTABILITY”

The performance gap in America between the nation’s most and least proficient students is larger than in all but a few member countries of the Organization for Economic Cooperation and Development (OECD). Yet despite this shameful gap, school accountability has developed something of a bad name in the United States, in part because NCLB’s system of escalating sanctions for failing schools has rarely had a positive counterpart in the form of incentives and rewards for schools doing a superior job of closing the achievement gap. NCLB’s many sticks (and rarely utilized carrots), buttressed by the resistance of some educators to accountability, has meant that the positive power of accountability to boost learning is often overlooked. As Secretary of Education Arne Duncan remarked at a 2009 EEP forum, K-12 education is “one of the only industries where we are scared to talk about excellence and are scared to talk about success—and that has to change.”

Successful practitioners of accountability share the Education Equality Project’s tenet that holding teachers, principals, and district officials accountable for student achievement is vital to closing the achievement gap. At high-performing inner-city schools like the KIPP academies, Achievement First charter schools, the Green Dot schools, and Geoffrey Canada’s Promise

Academy in Harlem, frequent testing and accountability for results help propel mid-course corrections in instruction and enables teachers to tailor instruction to the specific needs of individual students. The regular use of assessments also helps principals to identify and reward their most effective teachers, pinpoints weaknesses in less effective instructors, and empowers schools to rid themselves of the small minority of poor instructors who fail to improve even after receiving assistance. “Our best teachers today are using real time data in ways that would have been unimaginable just five years ago,” says Secretary Duncan. “They need to know how well their students are performing. They want to know exactly what they need to do to teach and how to teach it. It makes their job easier and ultimately much more rewarding.”

“Accountability” at these gap-closing schools is not an enemy but a friend. And the record of these schools underscores historian Ravitch’s observation that tests and standards “play a constructive role. They tell public officials whether new school programs are making a difference and where new investments are likely to pay off. They tell teachers what their students have learned—and have not. They tell parents how their children are doing compared with others their age.”

In fact, every educator and every parent ought to *want* schools to be in a position to answer such elemental questions about schools and student performance—the fact that most schools still cannot do so is a disgrace. As the April 2009 report “Smart Options” from the Coalition for Student Achievement highlights, accountability advocates are not questing for the obscure. A good longitudinal data system, the coalition report notes, would simply provide critical but basic information about student achievement. With accountability data, the report states, “parents will know in real time whether their child is on track to graduate from high school prepared for college and career success and how they can help. Teachers will know whether each and every student is on track for college and careers and, if not, how they can more appropriately tailor their instruction for individual students. Principals will be able to perform more accurate evaluations and customize teachers’ professional development plans. . . Districts will know where their best teachers are, where they came from and where they go, and which investments make the most difference for students.”

Thankfully, urban school districts, state education agencies, and the federal government have started to fund, install, and utilize the performance measures and sophisticated longitudinal data systems that can make accountability a potent force for improvement in all schools. The American Recovery and Reinvestment Act set aside a minimum of \$250 million for the purpose of developing state data systems, and both President Obama and Secretary Duncan have spoken bluntly of the need for better data-gathering and accountability for results by states and districts. “The time for finger pointing is over—the time for holding ourselves accountable is here,” President Obama declared in his education speech to the U.S. Hispanic Chamber of Commerce in March 2009. Longitudinal data systems that track student progress, the president said, “tell us which students had which teachers so we can assess what’s working and what’s

not. That's why we're making a major investment in this area that will cultivate a new culture of accountability in America's schools."

Given the historic opposition of educators to accountability, the rapid adoption in the last five years by states and districts of longitudinal data systems that can track student progress over time is encouraging. In 2005, just 21 states could determine which schools produced the strongest academic growth for their students, but 39 states can do so today, according to the Data Quality Campaign (DQC). Still, states and districts have a long way to go before most states will have robust data systems for holding educators accountable for student outcomes. Citing DQC figures, Secretary Duncan notes that only six states (Alabama, Arkansas, Delaware, Florida, Louisiana, and Utah) have "comprehensive data systems meeting the basic elements of a good system." "Far too few states," as President Obama has said, "have data systems like the one in Florida that keep track of a student's education from childhood through college. And far too few districts are emulating the example of Houston and Long Beach, and using data to track how much progress a student is making and where that student is struggling. That's a resource that can help us improve student achievement."

To date, the biggest shortcoming of state data systems is that most states still are unable either to link teacher performance to student performance or to provide teachers with student data that can help improve classroom instruction. "Even the best data collection system is worthless if it does not change what goes on in the classroom," observes Pennsylvania Governor Ed Rendell, chairman of the National Governors Association. Yet fewer than half of all states currently have a teacher identifier system that enables administrators to match teachers to students. And according to a 2009 study by SRI International for the U.S. Department of Education, just two in five teachers can access their current students' scores on benchmark or diagnostic tests. A mere 11 percent of teachers can access multiple years of standardized test scores for individual students. If teachers in inner-city schools were not flying half-blind, their ability to aid struggling students, tailor instruction, and narrow the achievement gap would be bolstered. In urban classrooms, knowledge is power but ignorance is not bliss.

THE LESSONS OF NO CHILD LEFT BEHIND

The standards movement of the late 1990s, which culminated in the passage of NCLB in 2002, created a long-overdue national focus on racial and ethnic achievement gaps and raising the test scores of the lowest-performing students. As Secretary Duncan has observed, there is a lot not to "like about No Child Left Behind but I will always give it credit for exposing our nation's dreadful achievement gaps. It changed American education forever and forced us to take responsibility for every single child, regardless of race, background, or ability." For all its flaws, NCLB imposed a uniform school accountability framework for the first time in the country's history and created a series of progressive interventions that led to the eventual closure or reconstitution of the worst performing schools in each state.

The best studies of NCLB's impact suggest that the law helped raise the math and reading scores of low-performing students in elementary and middle schools (but not in high school) and helped to modestly narrow the achievement gap, though existing studies can only demonstrate a correlation with changes in student achievement, not causation. A study, for example, by the Thomas Fordham Institute of math and reading performance of students in the lowest achieving decile found significant gains on the NAEP by the lowest-performing students from 2000 to 2007, especially in math. The Fordham study also concluded that the performance of students in the lowest decile in reading and math rose faster during the NCLB-era than in the 1990s. Similarly, a U.S. Dept. of Education study of the Title I program found that students in high-poverty schools made larger gains in reading and math on the NAEP than students in low-poverty schools following the passage of NCLB.

One beneficial, unambiguous impact of NCLB and its accountability provisions is that states and districts are now able to systematically identify their worst schools for the first time. Under NCLB's regimen, schools that fail to make Adequate Yearly Progress (AYP) five years in a row are up for reconstitution and a range of potentially far-reaching reforms. "Because of the state and federal assessment data we now have," writes EEP signatory and former Secretary of Education Margaret Spellings, "we know precisely which districts, schools, teachers and students need help and which are doing well. We've diagnosed the problem—the approximately 4,000 schools that have failed to meet their annual goals for five straight years and the 2,000 high schools that produce more than half of all dropouts."

Yet while NCLB appears, for the most part, to have had a positive impact on the performance of low-income students, its accountability provisions have created a number of unintended consequences as well, several of which have not benefited disadvantaged students. The law set a laudable if utopian goal of having every student in grades three through eight be proficient on state tests in reading and math by 2013-14 but it let states define "proficient." Most states responded to the law's ambitious mandate by dumbing down standards to ensure that the vast majority of students would be deemed proficient. As Secretary Duncan pointed out at a 2009 EEP-sponsored forum, "in too many places where [students] are 'meeting standards', you are barely prepared to graduate from high school and absolutely inadequately prepared to go to a competitive university, let alone graduate. And the only way we can reverse that is to stop dumbing down 56 standards in 56 states [and territories]." Three years after the enactment of NCLB, the U.S. Department of Education piloted an AYP model in a small number of states that added some measures of student growth to the existing proficiency model. But in most districts all that has mattered for principals, administrators, and teachers was whether their schools met a minimum, state-defined level of proficiency.

Not surprisingly, setting 50 different goalposts for student achievement created glaring inequities from state to state in NCLB's accountability system. As Randi Weingarten, head of the American Federation of Teachers pointed out, "imagine the outrage if, during the Super Bowl,

one football team had to move the ball the full 10 yards for a first down while the other team only had to go seven. Imagine if this scenario was sanctioned by the National Football League. Such a system would be unfair and preposterous.” Reports from independent research groups with titles like “The Pangloss Index” and “The Accountability Illusion” documented in painful detail how the dumbing down of standards by states led to lousy schools receiving state stamps of approval. The Education Sector’s “Pangloss Index”, for example, reported that in Birmingham, Alabama—where less than 40 percent of students graduated on time and which had some of the lowest test scores in Alabama—just five of the district’s 65 schools were deemed to be “in need of improvement.”

Other studies documented a slew of unintended consequences and strategies that states used to “game” NCLB’s accountability rules. As *New York Times*’ columnist David Brooks summarized earlier this year, “as our ability to get data has improved, the education establishment’s ability to evade the consequences of data has improved, too.” Some districts focused instruction on the “bubble kids”—students who were just below the proficiency bar—while largely ignoring advanced students or students who were far behind and might not be able to make the proficiency cutoff. Other districts narrowed the curriculum and “taught to the test” in math and reading, two subjects that schools were required to test students under NCLB. Yet others tried to manipulate test results by excluding special education students or adjusting the size of the subgroup of students on tests that could be counted toward a school making AYP.

These strategies to avoid accountability reaffirmed the validity of Campbell’s Law, which holds that “the more any quantitative social indicator is used for social decision making, the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it is intended to monitor.” Still, it is easy to exaggerate the prevalence and significance of states and districts trying to game NCLB’s accountability system. The most comprehensive 50-state analysis of test scores to date under NCLB, released by the Center on Education Policy in June 2009, failed to find that test scores rose chiefly among “bubble kids” but not among advanced students and the lowest-performing pupils. Similarly, the 2008 NAEP for music and the visual arts failed to document a widely anticipated narrowing of the music and arts curriculum following the passage of NCLB. In 2008, eighth graders received roughly the same amount of music and visual arts instruction as in 1997, the last time the NAEP music and visual arts assessment was administered.

Many critics have been too quick to flatly dismiss both NCLB and the uses of accountability and testing in school reform. The Education Equality Projects rejects the argument that all accountability measures are so subject to manipulation that they are doomed to fail or that multiple choice fill-in-the-bubble tests used by most states to meet NCLB’s testing requirements are meaningless measures of student achievement. As E.D. Hirsch Jr., a critic of current reading tests wrote earlier this year in the *New York Times*, the “much maligned, fill-in-the-bubble reading tests are technically among the most reliable and valid tests available. . . . we do not

need to abandon either the principle of accountability or the fill-in-the-bubble format. Rather we need to move from teaching to the test to tests that are worthy teaching to.”

Much the same point might be made about NCLB’s flaws. The chief lesson that educators should draw from NCLB’s implementation is not that accountability should be abandoned but rather than changes need to be made to NCLB to make for a better accountability system. As the foregoing discussion suggests, three fundamental reforms to NCLB hold obvious promise:

- 1) Moving toward a system of national standards to curb the race to the bottom by states;
- 2) Shifting to a value-added system for accountability that primarily tracks growth in student learning rather than relying solely on the attainment of minimum proficiency levels; and
- 3) Creating a better differentiated system of school improvement sanctions so that schools that narrowly miss AYP with one subgroup of students are not lumped together with abysmal and persistently failing schools.

Still, as lawmakers and educators move to modify NCLB’s accountability rules in the coming year, they should not lose sight of the fact that existing accountability schemes are already showing promise of narrowing the achievement gap. As noted, math and English test scores have risen since the introduction of NCLB for students in high-poverty schools and among the lowest decile of students. But it is also the case that the unintended consequences of government policy can be positive as well as negative—a sticking point that economists and researchers often overlook.

To paraphrase songwriter Johnny Mercer, not all analyses of school accountability fail to accentuate the positive. A 2005 analysis of the impact of the 1990s’ standards movement by Eric Hanushek and Margaret Raymond found that state accountability systems with consequences for poor performance did boost NAEP math and reading scores. “The movement toward stronger accountability in schools has also suggested too many that there would be adverse consequences—more exclusions, higher dropout rates, a narrowing of the curriculum, and the like,” Hanushek and Raymond noted. “We conclude that the negative impacts are likely to be considerably overstated.” Even Diane Ravitch, an NCLB critic, believes that the state standards movement was a boon to disadvantaged students. Writing in *Time* in 2000 she asserted that “in the past few years, we have seen the enormous benefits that flow to disadvantaged students because of the information provided by state tests. Those who fall behind are now getting extra instruction in after-school lessons and summer programs. In their efforts to improve student performance, states are increasing teachers’ salaries, testing new teachers and insisting on better teaching education.”

More recently, studies of accountability measures have provided fresh evidence that accountability requirements may boost student performance in failing schools. Surprisingly few researchers have actually studied the school-level impact of accountability. But two recent National Bureau of Economic Research (NBER) analyses of Florida and New York City's school report card grading programs found that receiving a grade of an "F" propelled schools to rapidly institute reforms that boosted math and reading test scores. In both locales, the NBER analyses found suggestive evidence that the bump in test scores resulted from real improvements in instruction, such as an increase in after-school tutoring, in-school supplemental instruction, extended schools days and school years, summer school, mentors, and more direct instruction, rather than stemming from an artifact of gaming.

To be sure, no school accountability system is flawless. But educators must not let the perfect become the enemy of the good, as so often happens in our nation's inner-city schools. "To throw up our hands and say 'But we cannot measure performance in social sectors the way we do in business' is simply lack of discipline," Jim Collins writes in *Good to Great and the Social Sectors*. "All indicators are flawed, whether qualitative or quantitative. Test scores are flawed, mammograms are flawed, crime data are flawed, customer service data are flawed, and patient-outcome data are flawed. What matters is not finding the perfect indicator, but settling upon a consistent and intelligent method of assessing your output results, and then tracking your trajectory with rigor." Secretary Duncan makes a similar point, albeit employing a different analogy. "To somehow suggest that we should not link student achievement and teacher effectiveness is like suggesting that we judge a sports team without looking at the box score," Duncan notes. "It's like saying 'since standardized tests are not perfect, eliminate testing until they are [perfect]'. I think that's simply ridiculous. We need to monitor progress. We need to know what is and is not working and why."

Someone once observed to Voltaire that "life is hard"—to which Voltaire retorted, "compared to what?" It is the compared-to-what test that policymakers must constantly keep in sight when assessing reforms to school accountability systems. Al Shanker, for example, was concerned that performance-pay schemes for teachers had to be done "very carefully" or could turn "dysfunctional." Yet he was optimistic that performance-pay programs would ultimately improve student outcomes. "I'm sure that we can develop such a system and that it would be pretty good," Shanker said in 1993. "Its flaws would be very small compared to what we have now—or what you would have without such a system."

FIVE GUIDING PRINCIPLES FOR REFORM

The Education Equality Project’s recommendations for improving accountability and using it as a force to narrow the achievement gap are animated by five guiding principles. **The overarching goal of a strong accountability system is to ensure that all students graduate from high school, prepared for a career or to enter college without requiring remediation.** The EEP’s five principles for achieving such a system are:

- 1) Educators should be measuring outcomes, not inputs or compliance.** The problem in schools today is not that they gather too little data, but rather that they are obligated to collect mountains of data that have little or no bearing on raising student achievement. A recent benchmarking survey by the Data Quality Campaign of 69 school districts found that district administrators were “overwhelmed by the sheer number of data collects from the state for compliance reporting.” One superintendent reported that his district had 83 state data collects a year but that just two percent of the data could be used for improving student achievement. As Susan Neuman stated in *Changing the Odds for Children at Risk*, in education “results are defined in terms of measurable improvements in children’s achievement—nothing less will do.”
- 2) A good accountability system monitors academic progress and performance—it measures, that is, what schools contribute to students, not what students bring to schools.** As noted, high-functioning accountability systems provide a longitudinal look at student growth over time, not just a snapshot. Shifting from a proficiency-based system to one that relies primarily on value-added assessments of student growth for evaluating school performance would thus reduce the pitfalls of NCLB. Secretary Duncan put it well when he said that he is looking for “demonstrable gains in student achievement, where we see students performing better than they have done in the past . . . You can go beyond [considering just test scores]. You can look at AP results; you can look at teacher attendance, student attendance, graduation rates. You can look at a reduction in violence. There is a range of indicators. But at the end of the day, it’s about student achievement.”
- 3) Accountability should entail consequences—both sanctions for poor performance and rewards for outstanding performance in closing the achievement gap. An effective accountability system also provides ongoing feedback, such as early warning alerts and progress reports that enable district officials, principals, and teachers to improve classroom instruction and boost student learning. The point of gathering accountability data is not to collect it but to use it to advance achievement.** At the start of the standards movement in the 1990s, a number of states experimented with report card-style accountability, where states and districts reported school test scores but attached no consequences to either poor performance or outstanding performance.

At the time, Al Shanker was one of the few educators who opposed toothless report card accountability. “I would prefer a world where you didn’t have to force people to do things—where they would pick up a book by Shakespeare and say, ‘Gee, I’d love to read this,’” Shanker observed. “Unfortunately, the vast majority of young people won’t do this—unless they have to. . . None of these proposed reforms is going to work in any large way unless there are stakes. Stakes change everything.”

Accountability, in short, if it is to mean anything, should not just entail building a better data-mousetrap but rather those teachers, principals, and districts bear some responsibility for the performance of their students. Persistently dreadful schools that fail to improve even after receiving assistance need to be closed—not studied or put on the agenda for another PTA meeting.

Hanushek and Raymond’s 2005 study referenced earlier found that report card-style accountability in fact failed to raise student achievement appreciably but consequential accountability systems did boost NAEP math and English scores. Unlike NCLB, which primarily imposes sanctions on failing schools, a productive accountability system also rewards schools when students show strong academic growth—providing a stronger incentive for educators to welcome performance assessments and raise minority achievement. These incentives could come in the form of performance pay to principals and teachers, bonuses for effective teachers who work in high-poverty schools or with underserved populations, and payments to low-income students who show especially strong academic growth.

- 4) The race to the bottom in state standards has weakened NCLB-style accountability. To make accountability more effective and kick-start a “race to the top,” states should sign on to the growing voluntary movement to create rigorous career and college-ready, internationally benchmarked national standards.** A low-income middle school student in Baltimore should not be held to different learning standards in algebra than a middle school student in affluent Brookline—or Beijing for that matter. Yet as Secretary Duncan has observed, up until 2009, national academic standards were “unimaginable” in the United States. No more. As of June 1, 2009, the governors and chief state school officers of 46 states—accounting for 80 percent of the nation’s K-12 population—have signed a memorandum of agreement committing their states to the development of voluntary, grade-by-grade common standards in math and English Language Arts. It remains to be seen yet whether these common standards will truly be rigorous, and whether states with low academic standards will be willing to embrace rigorous national standards, even though the higher standards will spotlight the poor quality of local schools. Still, the new NGA/Council of Chief State School Officers’ Common Core State Standards Initiative marks the beginning of an important sea change in K-12 education.

- 5) A high-functioning accountability system is transparent, user-friendly, and collects comparable data across schools on academic performance and attainment that is not subject to easy corruption. A well-focused accountability system also primarily assesses academic performance in core subjects—rather than holding schools accountable for student health habits, art appreciation, community involvement, cooperative behavior, social skills, good citizenship, and electives like theatre and computer programming.**

If district officials, principals, teachers, and parents are to use data gathered for accountability purposes to advance student learning, the data should come in a form of progress reports and other updates that are consistent, transparent, replicable, and user-friendly. If it takes a Ph.D in statistics to figure out a value-added analysis of a student's progress, a teacher assessment, or a fourth grade classroom's math scores, that assessment, no matter how sophisticated, is a failure. "Part of the problem is that people [and parents] don't know how to read data, how to sift through it or understand it and that's really a challenge for all of us," say Secretary Duncan. "We cannot communicate an undecipherable code." To help ensure that accountability data are transparent, it is vital that state and districts rely on common formulas, and a central database where applicable, to define outcome indicators. A parent might assume, for example, that when schools report high school graduation and dropout rates, they are all reporting numbers calculated according to the same formulas. Due to recent NCLB regulations issued in December 2008, graduation rate calculations will soon be the same in most schools around the country. But that is not the case today.

Schools have many legitimate purposes besides advancing student learning, including cultivating good citizenship, developing an appreciation of the arts, and discouraging students from unhealthy habits, like smoking and eating fatty foods. But while school leaders may feel that some or all of these goals are an essential part of a student's education, the Education Equality Project does not believe that educators should be held accountable for student performance in these areas. Measuring student and school performance on any of these dimensions is notoriously subjective—and far more prone to statistical gaming than analyses of math and English Language Arts scores. Schools should educate students about the perils of smoking and poor nutrition. But schools cannot truly be held accountable if Johnny starts smoking or Janey becomes obese. Not all educational ends are equal—and a good accountability system cannot be all things to all educators. Indeed, such diffuse accountability would undermine educators striving to close the achievement gap. As the old adage puts it, "anything is possible, but everything is not."

FIVE POLICY RECOMMENDATIONS

Drawing on the preceding principles, the Educational Equality Project endorses five policy recommendations for improving K-12 school accountability. They are:

- 1) The EEP staunchly supports the movement toward voluntary national academic standards. But it cautions that many obstacles lie ahead over the next several years before a common core of standards will assist accountability efforts and reduce the achievement gap. The issue is not merely reaching agreement on the contentious issue of a rigorous common core of standards in math and English Language Arts. Curriculum and assessments must also be developed to make the standards useful for accountability purposes. Accomplishing all this is doubly demanding given the fact that these curricula and assessments are not to be aligned anymore with graduating from high school, but rather with having students be career and college-ready, and prepared to compete with their peers in high-achieving countries. The Common Core State Standards Initiative, the NAEP, and education nonprofits should be readying themselves to develop these critically important assessments and curricula. As this effort proceeds, the Education Equality Project also supports other initiatives to move toward voluntary, common college-ready standards, such as Achieve's American Diploma Project and multi-state consortiums working to develop common standards, perhaps with support from the new federal Race to the Top fund.*

The evolving consensus in favor of voluntary but rigorous national standards could eventually make it much easier to design high-performing accountability systems by effectively ending the race to the bottom in states that set low academic standards. For the first time, a rare confluence of factors may mean that the grade-by-grade development of a common core of standards in math and English Language Arts is more likely than not. It is politically fortuitous that the current national standards movement is driven by states, led by the chief state school officers and the nation's governors, rather than being imposed top-down from the federal level. While many states have pushed for low standards in the past, nationalization of standards would not get far at present without state support. Both of the leading teacher unions, and powerful nonprofits, including Achieve, the Gates Foundation, the ACT, and the College Board, have signed on to the state-led effort, creating an unprecedented coalition for change.

The Memorandum of Agreement that 46 chief state education officers and governors recently signed to develop a voluntary common core of standards stipulates that no state will see a decrease in the level of student expectations from current state standards. In fact, the Common Core State Standards Initiative is expected to develop far more rigorous and relevant standards than most state standards, if only because existing standards are geared to high school graduation, not to preparing students to be college-ready in a global economy. Assuming that the NGA and Council of Chief Schools Officers (CCSSO) succeed in crafting a workable consensus

on the content of national math and English Language Arts standards, the need to develop curriculum and assessments aligned to the new standards remains pressing. Despite the challenging balancing act that the NGA-CCSSO campaign for national standards faces, the Education Equality Project believes that the Common Core State Standards Initiative, the NAEP, and leading nonprofits must begin to lay the groundwork for developing curriculum and assessments that will make new national standards useful accountability tools for elevating student achievement. The EEP was heartened by Secretary Duncan's announcement in June that the federal government will be setting aside \$350 million of the stimulus funds to develop assessments to test if students are meeting the new achievement benchmarks.

- 2) *Apart from developing national standards, much remains to be done to create true accountability in our nation's schools. The most important step in the near-term would be to rewrite NCLB during its reauthorization to shift from the current proficiency-based model for assessing academic achievement to a model that relied primarily on value-added assessments of student growth. To be sure, a value-added centric model should include some proficiency-based indicators and other measures of academic attainment as well. But the most serious school sanctions, including the requirement to reconstitute a school, should be reserved for schools where students both perform at low levels and fail to make progress compared to their peers. When reauthorized, NCLB should be rewritten to do a better job of differentiating school improvement needs.*

It is no secret that NCLB's proficiency-based model for assessing student achievement drove many states to define down standards. Proficiency cut scores and other aspects of the AYP calculation led at least some districts and schools to teach to the "bubble kids," narrow their curriculum, and adopt various gaming strategies to avoid accountability. A value-added model, which compares academic growth among students and schools to other students and schools with similar demographics and background characteristics, is a fairer way of assessing school performance since it measures what schools contribute to students, not the skills that students bring to school. (It is also less subject to corruption because it does not utilize a proficiency cut score but instead measures growth). In practice, the outcomes of these two accountability methodologies vary enormously, as Thomas Toch has pointed out in *Education Week*. In Dallas, one of the few districts to rate schools using both a NCLB proficiency-based metric and a value-added methodology, researchers found that "schools ranked 94th, 77th, 83rd, and 107th among the city's 206 schools under NCLB placed second, fifth, eight, and 16th under the city's value-added ratings."

The Education Equality Project would like to see more districts and states implement accountability systems that depend primarily on value-added metrics, but supplement the data with proficiency-based metrics. In New York City, for example, 55 percent of a school's report card accountability grade is based on student progress. The progress score consists of the percentage change in individual students' math/English scores, the average change in

proficiency among all students in the school, and the average change in the proficiency of the lowest third of students. Underscoring the importance of raising achievement among lower-performing students, schools can also receive special credit for exemplary student progress among subgroups of Hispanics, black students, English language learners, or special education students. On a school peer index, school performance is then compared to the 20 schools ranked just above the school and the 20 schools ranked just below it, based on the composition of the peer schools' student bodies or the performance of students on state exams prior to their arrival at their schools.

However, accountability grades for schools are not limited in New York City to value-added metrics. Thirty percent of a school grade is based on student performance (i.e., the percent of students who are proficient on state tests), with the remaining 15 percent of the overall score accounted for by attendance and evaluations of the school environment from annual surveys of students, parents, and teachers. Pre-NCLB, Florida had established a similar accountability system that also combines value-added and proficiency ratings, and provides special credit for learning gains among low-performing students.

Value-added metrics are more complicated to calculate and have bigger margins of error than proficiency-based metrics. That's one reason why sanctions and rewards in a value-added system for schools, teachers, and principals should primarily be reserved for opposite ends of the bell curve—say, the bottom quintile of performers and the top quintile—rather than meted out based on fine-grained statistical differences of dubious value. At present, the AYP regime for monitoring school improvement in NCLB fails to distinguish adequately between schools that are struggling with some subgroups of students and schools that fail their students school-wide, year after year.

The most serious sanctions ideally should be reserved for schools that abjectly fail both to promote student growth and have low levels of student achievement, while the biggest bonuses should be reserved for schools that produce both large learning gains and have high levels of achievement. No school, however, should be able to earn a passing grade under a value-added system if it achieved less than a year's worth of learning in each grade assessed.

- 3) *Contrary to some anti-testing rhetoric, a well-designed accountability system does not rely on test scores in math and English Language Arts as its sole outcome indicators—particularly once students enter high school, where college and career-readiness take on added importance. Among other outcomes that should be monitored for accountability purposes, districts and high schools should be tracking attendance, discipline data, Advanced Placement and International Baccalaureate participation, enrollment in a college-prep curriculum and honors courses, on-time high school completion, SAT and ACT scores, job entry and/or college enrollment, post-high school participation in remedial postsecondary courses, and years of college completed.*

Even in elementary and middle school, educators ought to monitor—and be held in accountable for—student performance in core subjects other than math and English Language Arts through the use of state assessments and other tests in science, history, foreign language, and writing. However, the EEP does not support formally holding schools and teachers accountable for student performance in electives like music, visual arts, computer programming, dance, home economics courses, and theatre. It is also opposed to holding schools formally accountable for the mental health of students, student social skills and problem resolution skills, success in promoting citizenship, and the coordination and provision of early childhood education, out-of-school services, summer learning opportunities, parent support services, and social services in the community.

Many state accountability systems are driven by student scores on math and English Language Arts tests in grades three through eight because NCLB requires assessments during those years. Test scores, including results from multiple choice assessments, are vital indicators and should figure prominently in any accountability system. But to avoid narrowing the curriculum, states and districts ought to also be using state assessments and other tests to measure performance in other core subjects, like science, history, and foreign languages.

At the same time, formal accountability systems are not necessary for elective subjects like music or computer programming. It is perfectly appropriate, of course, to assess student knowledge in electives. The NAEP, for example, periodically performs special NAEP art assessments. In 1997, the NAEP asked eighth graders, among other questions, to identify a collage by Romare Bearden as “impressionism,” “photographic realism,” “surrealism,” or “semi-abstract representation”; it also asked students to identify both time signatures in written music and the composers of classical pieces like the “March” from Tchaikovsky’s Nutcracker Suite. There is nothing wrong, and there is much to be gained, by assessing the state of student knowledge in the arts. But urban schools should not be subject to sanctions or bonuses in formal accountability systems if their eighth graders do not know the difference between impressionism and semi-abstract representation, or if they fail to identify the composers of classical pieces.

In much the same vein, the EEP opposes the efforts of some advocates to wholesale expand school assessments to hold schools formally accountable for student social skills, mental health, problem resolution skills, success in citizenship, and the satisfactory coordination and/or provision of early childhood education, and out-of-school and summer learning opportunities. While many of these same advocates criticize both the quality and utility of current math and reading assessments in state accountability systems, they are curiously blithe about the ability of states and districts to create a multi-billion dollar system of trained inspectors—who would

be responsible for equitably assessing the nation's 95,000 schools on a regular basis on nearly every dimension of school performance imaginable, no matter how ill-defined.

When students reach high school, accountability systems should shift from concentrating heavily on test scores to include more measures of educational attainment: Did the student graduate on time? Did he or she enroll in college? How many years of college did the student complete? Few elementary or middle school students drop out of school. But even high test scores don't help a disadvantaged student much if he or she drops out of high school, or abandons college within the first few months. Congressman George Miller, chair of the House Education Committee, has proposed that high school graduation rates be included with test scores in evaluating schools, and the EEP supports as well adding other non-test score measures of readiness and attainment to accountability systems for high school students. Louisiana, for example, has developed a model Dropout Early Warning System that tracks student attendance, grade point average, discipline data, and student age so schools can identify and work with students at risk for dropping out. In Chicago, former superintendent Arne Duncan created an "on-track indicator" for high school freshmen after the Consortium on Chicago School Research discovered that ninth grade attendance and course grades were even more important predictors of high school graduation than eighth grade test scores and family background.

Monitoring attainment and readiness indicators in high school is doubly important because of the new emphasis in the standards and accountability movement on preparing students to be career and college-ready, according to internationally benchmarked standards. Traditionally, high school graduation has marked the final certification of readiness under state standards. But in today's global economy, a high school diploma alone is rarely enough to land a decent job. President Obama has set a goal that every student should have at least one year of post-high school training. Unfortunately, colleges and universities report that high school graduates often arrive far from college-ready and must enroll in remedial classes as soon as they set foot on campus. The need for remedial education is particularly acute among low-income and minority students.

High school educators, in short, face a host of new accountability responsibilities as they fashion a better pipeline for the college freshmen of the future. What percent of students are enrolled in a college-prep curriculum—and do all students have equitable access to AP and college-prep courses? Which high school courses and skills are most helpful to students when they get to college? What percent of graduates are required to take remedial courses when they enter a postsecondary institution? These, and many other indicators, will eventually help build a longitudinal accountability system that is critical to closing the achievement gap. At the moment, however, the development of college-readiness standards and data systems is still very much a work in progress. A June 2009 survey of the 50 states and the District of Columbia

by the Editorial Projects in Education Research Center found that for the class of 2009, just 20 states had defined the skills and knowledge necessary to succeed in entry-level college courses.

Elevating existing standards for high school students to new, rigorous criterion for college and career readiness is an ambitious task. Yet it can be done—as the state of Florida has shown. For each high school in the state, Florida puts together a High School Feedback Report that details how students perform on admission tests and how they subsequently fared in college courses. Florida’s database, as *Education Week* recently reported, is “available online to anyone, allows users to see, at the school level, the percentage of students who earned a diploma or [GED] certificate in a given year. It also provides a look at the percentage of students who took dual-enrollment college courses while in high school and tracks those who took the highest-level high school courses in mathematics and science. Because the data system is linked with Florida’s postsecondary institutions, high schools find out the percent of their students who must take remedial courses, the percentage that took and passed freshman math and English courses, and the percentage who maintained at least a 2.0 GPA as their college careers continued.” In each of these cases, school-level numbers appear side-by-side with district and statewide numbers, giving parents and students a clear measuring stick for gauging their school’s performance.

Florida also generates a Performance on Common Placement Tests report for high schools that reports the percentage of students who scored at or above the college-level “cut-score” on the SAT, ACT, and Florida’s community-college placement tests. (The cut-scores determine if students can enter credit-bearing English and math courses at Florida colleges and universities.) By establishing a true longitudinal, pre-K-16 data system, Florida—unlike most states—is now able to document the correlation between student scores on assessments and placement tests and, say, the need for college remediation. Every state should be able to do the same.

- 4) *While the Education Equality Project believes that multiple choice assessments provide valuable and valid information on student achievement, the current generation of state assessments should be more challenging and do a better job of assessing critical thinking and reasoning, especially among high school students. The EEP is all in favor of better assessments. In lieu of states shifting away from multiple choice assessments in high school, the EEP believes that national tests like the NAEP, SAT, and ACT can provide important data on student performance and higher-order thinking skills that may be missing from existing state multiple-choice assessments. Meanwhile, states should also move toward doing more of their testing online and incorporate online adaptive assessments for instructional improvement. The EEP supports broader and richer performance assessments. But it does not favor switching to portfolio assessments in state accountability systems in place of tests.*

Multiple choice assessments have become the *de facto* test of choice during the last two decades because they are relatively inexpensive to administer and score and because they have a well-documented validity. Despite their shortcomings, “fill-in-the-bubble,” multiple choice assessments have provided teachers, parents, and principals with invaluable information that they previously lacked on student cognitive abilities and the dimensions of the achievement gap.

Today, as the aspirations of K-12 educators expand to include making students college-ready, the time has come for state tests and exit exams to do a better job of assessing higher-order thinking, reading, and writing skills, especially among high school students. There are several pathways to creating better assessments. States could retain the multiple choice format of most of their assessments but make multiple choice questions more demanding. States could begin to try to follow the lead of national tests like the NAEP, by adding numerous constructed response items that require students to work out factual or written answers on their own. Online testing, pioneered by states like Virginia, also holds great promise for incorporating adaptive assessments for instructional improvement.

Most high-achieving nations rely on tests dominated by constructed response items, which require students to apply knowledge and write more extensively. Many of these same nations also have national standards. At present in the United States, tests like the NAEP, SAT, and ACT often have limited utility in accountability systems because school curricula are aligned to state standards and not to national ones. But even if voluntary national standards do not take hold in the U.S., tests like the NAEP and SAT remain critical for assessing student performance because they allow educators to begin to compare student performance across states.

While the EEP supports richer performance assessments, it does not believe that states should abandon testing in favor of assessing teacher and student performance by evaluating and scoring student portfolios, exhibitions, and other non-test based methods for measuring content mastery. Portfolio-based assessments were a fad in some states in the late 1980s and early 1990s, but often proved costly, produced unreliable scores, and were difficult to incorporate into accountability systems. To cite one example, Bill Tucker notes in a recent Education Sector report that a 1992 RAND Corporation study of Vermont’s portfolio assessment program found “significant problems with the reliability of the program’s test scores.” Harvard professor Daniel Koretz, who authored the RAND report, explained that it was “difficult to make scores comparable in meaning from year-to-year and from school-to-school.”

- 5) *The Education Equality Project endorses the ten essential elements of longitudinal data systems (<http://www.dataqualitycampaign.org/survey/elements>) that the Data Quality Campaign has promoted and helped proliferate at the state level. **Good accountability systems, however, not only contain rich data but are transparent, clear, tailored to different users—and, above all—used to hold educators responsible for student***

outcomes. *It is for this last reason that the seemingly innocuous cause of building better data systems can be surprisingly controversial—and is so essential to closing the achievement gap. Reformers intent on narrowing the achievement gap can neither shirk the fight for better data nor minimize bad news about teacher performance and student achievement that new and better data may provide. Too often, however, accountability systems are still not well-linked to improving classroom instruction. The feedback instructional loop between district officials, principals and teachers must be faster, more effective, and part and parcel of school accountability systems.*

The Data Quality Campaign’s ten essential elements for state data systems, many of which have been cited previously, are entirely in keeping with the EEP’s goal of promoting longitudinal accountability systems that track student progress from pre-K to the conclusion of college. As the state of Florida has shown, the data that results from such a system should be intelligible to the interested layperson or parent, and it should result in distinct sets of progress reports—tailored to students, teachers, principals, and district needs.

Regular student and teacher progress reports, for example, would include diagnostic information on individual students that could be put to use by parents and teachers to provide assistance to struggling students. Early warning system reports help teachers, parents, and principals identify students at risk for dropping out and in need of additional assistance. Readiness reports identify whether students are on track for high school graduation and college. Value-added assessments of teacher performance help principals identify their most effective and least effective teachers. Feedback reports let principals and district administrators know how well schools and groups of students fare when they move on to the next level of secondary schooling or college. Peer school reports let principals and district officials see what schools are doing an excellent job of advancing student growth and which are failing to boost student learning. The list goes on.

While student data systems are growing exponentially, and while the data are being used in some school improvement efforts, the new accountability systems are typically “having little effect on teachers’ daily instructional decisions,” according to SRI International’s 2009 study for the U.S. Department of Education. “It is clear that in many districts, the use of locally generated data to inform instruction is an activity separate from uses of data systems containing student scores on standardized tests,” SRI International found. Districts and schools typically use standardized test scores both to respond to accountability concerns and to make sure that the curriculum is well-aligned with state assessments. By contrast, SRI International reported, “neither the type of assessment for which data are available nor the time frame of assessment activities serves the needs of classroom teachers making decisions on a daily basis.”

As Secretary Duncan pointed out in a recent speech at the Institute of Education Sciences conference, educators now have an enormous “opportunity for growth and progress [in the use of school data]. We have the money and we have the technology.” The final stumbling block, Duncan suggested, was not whether there was a way to better use accountability data but whether there was a will to use it. “The biggest barrier, the only remaining barrier in my mind,” Duncan said, “is do we have the courage? **It takes courage to expose our weaknesses in a truly transparent data system. It takes courage to admit our flaws and take steps to address them. It takes courage to always do the right thing by our children, but ultimately we all answer to the truth. You can dance around it for only so long. America’s children need your help.**”