
**“Value-added Assessment and Systemic Reform:
A Response to America’s Human Capital Development Challenge”**

Ted Hershberg
Professor, Public Policy and History
Executive Director, Operation Public Education
University of Pennsylvania

Prepared for the Opening Session
“Value-added Assessment and Accountability”

Aspen Institute’s Congressional Institute
*The Challenge of Education Reform:
Standards, Accountability, Resources and Policy*

Cancun, Mexico
February 22-27, 2005

“Value-added Assessment and Systemic Reform: A Response to America’s Human Capital Development Challenge”

Our nation, which has prevailed in conflict after conflict over several centuries, now faces a stark and sudden choice: adapt or perish. I'm not referring to the war against terrorism but to a war of skills -- one that America is at a risk of losing to India, China, and other emerging economies. And we're not at risk of losing it on factory floors or lab benches. It's happening every day, all across the country, in our public schools. Unless we transform those schools and do it now . . . it will soon be too late.

Louis Gerstner, former Chairman, IBM
Chairman, The Teaching Commission

Introduction: The Global Economic Context

Terrorism and the war in Iraq are high on the list of the nation’s concerns, but the greatest danger facing America, as Louis Gerstner recognized, is the challenge of human capital development. The nation’s public schools, the foundation for this effort, are still failing far too many of our children despite an investment of \$500 billion annually.

Sadly, we’ve known about this threat for quite some time. In 1983, *A Nation at Risk* suggested that if the mediocrity of our schools had been imposed by an unfriendly foreign power, “we might well have viewed it as an act of war.” We understood then that companies all over the world could buy fool-proof machinery that compensated for deficient worker skills and that billions of people were willing to use that machinery for a fraction of the wages American workers wanted. But recent trends and the highly visible discussion over off-shoring and outsourcing of high-end jobs make clear that globalization is about far more than displacing our nation’s blue-collar workers.

Every job in America is at risk. China has mandated that all of its college graduates will be proficient in English. China and India alone each year graduate 5.1 million from college and 400,000 engineers while our schools turn out 2.2 million college graduates and 60,000 engineers. We produce only two bachelor degrees for every 10 students who start high school. “The law of sheer numbers,” observes Hewlett-Packard’s CEO Carleton Fiorina, “is fairly compelling.”

The consequences for failing to meet the human capital development challenge are already severe. The last year the typical blue-collar worker earned enough for mom to stay at home and raise the kids was 1964. We maintained our standard of living despite the decline of real wages through these decades largely because women entered the labor force in record numbers and created two-income families, but that strategy has run its course. Virtually all gains in family income in the last 30 years have gone to the top fifth of American families, largely the result of new technologies that favor the better educated.

No Child Left Behind

Our last four presidents, the Congress, governors and corporate leaders came to understand that if America is to remain a stable, middle-class society, steps had to be taken to signifi-

cantly improve our system of public schools. Frustrated by precious little improvement in student achievement over the last two decades, the Congress in a bipartisan consensus in 2002 passed the *No Child Left Behind Act*, an unprecedented expansion of the role of the federal government in K-12 education.

NCLB broke new ground. It required states to set academic standards so we could move away from norms-based testing, which compared students to each other, and learn instead what students know and can do at given ages. It required states to test annually so we could have a basis for measuring changes in achievement levels. It forced schools to focus attention on the academic progress of long-neglected low-income and minority students, which is why many believe the law is one of the great civil rights victories in our history. And it introduced consequences for schools that failed.

The legislation was necessary and, despite some design flaws that will be addressed later, there must be no retreat from its goals. Nonetheless, it would be seriously wrong to conclude that NCLB represents a sufficient response to the human capital development threat posed by globalization.

NCLB was not and — given political realities at the federal level — could not in itself be a dramatic overhaul of the K-12 system. We remain the only developed nation where the federal government neither dictates nor closely coordinates the standards, assessments and curricula that are used in its schools. Yet the reason for breaking the long-standing “hands-off” role of the federal government was that to meet our human capital development challenge and to ensure the civil rights of children our schools had to do something never done before: educate *all* children, not simply the top fifth, and educate them to unprecedentedly high levels.

NCLB cannot directly produce fundamental change in American public education, but it can help the states transform our nation’s public schools. What follows is a description of the kind of system needed and a discussion of how we can work with NCLB to encourage the states, through incentives and regulatory changes, to promote the required changes so schools can graduate students who are able to use technology, think critically, solve problems and learn throughout their lifetimes. Whether this is possible to accomplish state-by-state remains to be seen, but at this moment, barring a sea change in the Congress, that is the best we can hope for.

America’s Public Schools: Past and Future

The system of public education now in place is largely unchanged from the nineteenth century when schools were set up to do three things. The first was to provide universal basic literacy, and America became the first nation in which virtually everyone in the labor force could read and write at the 6th grade level. The second was to socialize a highly diverse population — millions of immigrants from different nations, cultures and religions and even more millions of farmers who migrated to cities as American agriculture mechanized over the last 150 years — for success in an industrial economy. Students were taught to show up on time, respect authority, develop a work ethic and repeat monotonous tasks. Finally, using standardized tests and the bell-shaped curve, the schools identified and sorted out the top fifth of their students for higher education and the best and brightest among these went on to run the country.

The old system, concerned with quantity and cohorts rather than quality and individuals, was designed to let the cream float to the top. For the remaining 80 percent or so, there was little consequence because, for most of our history, they ventured forth into an industrial economy that provided ubiquitous jobs paying middle-class sustaining wages, but which required little in the way of advanced education or higher-level skills.

Our schools were enormously successful in these tasks, and there is no way to understand the emergence of America as an industrial superpower without acknowledging the key role they played. The problem is that people continue to behave as if the current school system — designed for a different century and a different economy — is the right one for the challenges ahead despite the record of the last three decades. Since 1970, notwithstanding an increase in real spending of over 100 percent per pupil, a decrease of 22 percent in the pupil-to-teacher ratio, and a doubling of the number of teachers with Masters degrees, student achievement has remained largely flat.

For schools to succeed in the 21st century, teachers and administrators must undergo the requisite education and training to master concepts and methods that were not necessary in the schools of an industrial economy. Educators will need to replace the “ability-based” notion that dominates the thinking of too many who work in our schools with an *effort-based* theory of learning: you are not simply born smart, but you can “get smart” with appropriate resources and high-quality instruction. They will need to see how teaching a *standards-driven curriculum* is very unlike using a bell-shaped curve to distinguish among students: all students must reach high standards. They will need to be proficient in using a *problem-solving pedagogy* in the classroom: memorization will always have a place in the learning process because no subject can be mastered without it, but it must never again serve as the dominant paradigm. They will have to learn how to *differentiate instruction*: “one-size-fits-all” cannot suffice in an era in which no child can be left behind and children have many different styles of learning. They must master *data-driven decision making*, a striking departure from an anecdotal approach that has long characterized our schools. Finally, they must learn how to shift from teacher-centered to *student-centered classrooms*: lectures serve the needs of some students, but everyone learns best when they bear more of the responsibility for learning.

But before our educators can master these new skills and knowledge, key problems that plague the profession of teaching must be addressed. According to the National Commission on Teaching and America’s Future: one in three teachers leaves in the first three years; 46 percent in the first five years; these rates are 50 percent higher in urban districts; and the profile of the “leavers” is stronger than the “stayers.” Of the 3.4 million current teachers, two million will leave in the next decade — three times as many through attrition as through retirement. To replace these teachers and to retain and attract the “best and the brightest” will require a transformed school system, one that will make teaching a more financially rewarding and intellectually satisfying experience.

A New System: Operation Public Education’s Comprehensive Reforms

The system necessary to encourage and support these changes will have to be governed through an entirely new set of rules and incentives. Many key elements of reform are already in place across the country addressing assessment, educator quality, compensation, professional development and capacity building. The challenge is to bring these together along with several

striking innovations so that the typical practice of piecemeal reform that produces marginal improvements can yield sustainable systemic change.

Operation Public Education, based at the University of Pennsylvania, has done this. It has developed new and comprehensive reforms that complement the federal law and provide tools that can transform America's schools. These reforms have been codified into legislative language and are now being promoted nationally with support from the Annenberg Foundation and the Carnegie Corporation. Winning support for these progressive and controversial reforms is a difficult, but not impossible task. The Superintendent of schools and the President of the Columbus, Ohio Education Association have committed to piloting the system and the two largest districts in Idaho — Meridian and Boise — have expressed a strong interest in forming joint district/union teams to implement the system in that state.

The OPE system provides new forms of educator evaluation that include *outputs* (student learning results) in addition to *inputs* (the observation of teachers in their classrooms). The compensation system enables outstanding teachers to earn higher salaries more quickly and is flexible enough to differentiate pay for those difficult-to-fill vacancies that arise from subject matter or less desirable working environments. It provides more fluid career opportunities so teachers can assume greater responsibility at earlier ages based on their effectiveness. Teachers needing remediation are required to undergo it, and ineffective teachers who are unable to improve must leave the profession. Professional development — opportunities to improve skill and obtain new knowledge — is substantially expanded so educators can continue to grow throughout their careers.

At the heart of the OPE system is an essential quid-pro-quo in which teachers accept accountability as individuals in return for a significantly expanded role in school management: (1) teachers are given responsibility for evaluating fellow educators through a peer-review process; (2) they play a key role in the remediation process that affects all key personnel decisions; and (3) because their success is now determined by what happens in their classrooms rather than at the bargaining table, they become full partners in the policy decisions that affect classrooms, such as professional development and curricula. While collective bargaining should continue, the OPE system dissolves the hard distinction between teachers as “labor” and administrators as “management” because the difficulty of systemic reform requires close cooperation among the key parties.

Finally, the mediocre, high-stakes standardized tests found in the large majority of states need to be replaced with a new integrated assessment system that would provide not only a high-quality “summative” exam at year's end focused on the development of higher-order thinking skills, but “formative” assessments throughout the school year designed to give teachers regular feedback in the form of suggested pedagogical interventions to support improved instruction for this year's students.

Value-Added Assessment and Accountability for Individual Educators

The OPE system rests on a foundation of individual-level accountability — both teachers *and* administrators are held responsible as individuals for student learning results. NCLB moved in the right direction in requiring accountability, but it fell short in making the school rather than individual educator the unit of accountability. Systemic changes of the type discussed above will

be achieved only when the careers of everyone working in our public schools are tied to successful learning outcomes.

But until now efforts to hold individual educators directly responsible for student learning — sometimes referred to as “pay-for-performance” or “merit pay” — have failed either because they were too subjective or relied on achievement scores, which are strongly influenced by family income (good jobs, years of schooling, positive attitudes about education, exposing one’s children to books and travel, and social and intellectual capital correlate highly with income) to differentiate effectiveness. The unfairness of such approaches was recognized and they have largely been abandoned.

At a recent national conference, Dan Fallon, Chair of Carnegie Corporation’s Education Division, explained the origins of the belief held by most Americans, lay people as well as K-12 educators, that the level of academic achievement is determined largely by factors beyond a school’s control. It can be traced back to James Coleman’s 1966 report — “only a small part of [student achievement] is the result of school factors, in contrast to family background differences between communities;” and the work of Christopher Jencks in 1972 — “the character of a school’s output depends largely on a single input, namely the characteristics of the entering children.” This understanding is reinforced for the public at-large when their metropolitan newspapers issue their annual “Report Card on the Schools,” revealing that wealthy communities almost always have the highest test scores.

“An implicit conclusion of the analyses put forward by Coleman and Jencks,” Fallon observed, is that “when it comes to student achievement, teaching doesn’t matter very much.”

But a spate of new studies now proves empirically that teaching matters enormously. To understand how this new conclusion can be arrived at, it is vital to grasp a fundamental distinction in the measurement of student learning. *Achievement* describes the absolute levels attained by students in their end-of-year tests. *Growth*, in contrast, describes the progress in test scores made over the school year. And here is the most important implication of this difference: high absolute scores on assessments such as the SAT are best predicted by family income. But if we are predicting student growth — progress made over the year — reports by education researchers Kain, Hanushek, Sanders and others demonstrate that *good instruction is 15-20 times more powerful than family background and income, race, gender, and other explanatory variables.*

Given the technological limitations of their era, Coleman and Jencks focused on what they could: achievement. But today, because researchers have access to data sets and technology that link the progress of individual students over time to the teachers who taught them, it is now possible to measure the impact of instruction on a student’s academic growth using a powerful new methodology called *value-added assessment*.

Value-added assessment is often confused with simple growth because the words themselves make it is easy to think about this growth as the “value” that is “added” over the last year. But the statistical method known as “value-added assessment,” as developed for the state of Tennessee by Dr. William Sanders when he was a professor of statistics at the state university, is a way of isolating the impact of instruction on student learning. Its great advantage is its ability to separate the annual academic growth of students into two parts: that which can be attributed to

the student and that which can be attributed to the classroom, school or district. Because individual students rather than cohorts are traced over time, each student serves as his or her own “baseline” or control, which removes virtually all of the influence of the unvarying characteristics of the student, such as race or socioeconomic factors.

Although there are several different value-added models in use today, only the Sanders model has been mandated for use statewide: in Tennessee since 1992 and most recently in Pennsylvania and Ohio as well as in over 300 other school districts in 21 states.

Under the value-added approach, test scores are projected for students and then compared to the scores they actually achieve at the end of the school year. Classroom scores that exceed projected values indicate effective instruction. Conversely, scores that are mostly below projections suggest that the instruction was ineffective.

At the same time this approach considers student factors such as the pattern of prior test scores, both those of the individual student as well as those of other students in the same class. If a student’s present performance is below projected scores, while students with comparable previous academic history in the same classes have done well, this is evidence of the student effect — external variables such as the home environment — which is outside the control of teachers and schools.

Since students’ projected scores are based only on their prior academic records rather than on race or socioeconomic background, value-added does not introduce bias: in other words, low-income children are not expected to do poorly and high-income students are not expected to do well. But because value-added traces the same students over time, thus accounting for family and neighborhood characteristics that so strongly bias absolute test scores, educators are not being penalized for circumstances beyond their control.

When value-added scores are collected for each classroom and averaged over three years, teachers have rich diagnostic information to improve their instruction and administrators have an empirical basis for evaluating teacher effectiveness. When these classroom scores are aggregated over entire buildings and districts, principals and superintendents can be held accountable for student learning results.

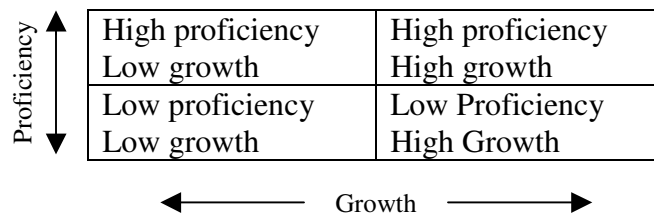
Value-added assessment is not without controversy given the complicated statistics on which it unavoidably rests. The RAND Corporation’s major study of value-added models identified important research questions about school effects, about the comparability of the instructional difficulty at different grade levels and subjects, and about the quality of the tests used, among others. Nonetheless, it concluded that: (1) the teacher effect is real; (2) it could be very large; and (3) it persists beyond the year in which it is first evident. Despite the problems associated with existing test-based accountability systems, the RAND researchers concluded that value-added models (VAM) “might actually provide less-biased and more precise assessments of teacher effects” than existing test-based systems and that as “policymakers evaluate alternative models for school or teacher accountability, VAM should be given serious consideration even in light of its limitations.”

In a balanced accountability system no educator should ever be evaluated solely on the basis of a single measure, not even one as powerful as value-added. Value-added scores, in the OPE system, for example, constitute half of an educator’s evaluation, while the other half is based on direct observation of performance using well-established protocols. Value-added models used for accountability should always be accompanied by safeguards, such as those developed by OPE, to ensure that educators are treated fairly as individuals. This is a moment when we must be willing to innovate, take risks and not let the perfect be the enemy of the good.

Some Problems with Adequate Yearly Progress

NCLB’s requirement that schools bring all their children to high standards by 2014 is a worthy goal. So, too, is the insistence that school-wide averages are not enough — student subgroups, including low-income, non-English-speaking, special needs and those of varied ethnicities, must meet these standards as well. The problem, however, is how to identify which schools are on target to meet these requirements. In most cases, NCLB’s AYP measures can correctly sort out successful schools from those that are failing their students. But for many schools, AYP measures do not provide a fair and complete assessment of school performance.

At the heart of this problem is the fact that AYP focuses on *achievement* to the exclusion of *growth*. The following chart helps us identify and understand AYP’s twin deficiencies. *Proficiency* (achievement), high and low, is tracked on the vertical axis, while *growth*, high and low, is tracked on the horizontal axis. In the bottom left cell are schools that are clearly not serving the needs of their students — providing them with low proficiency and low growth — and thus deserve to be sanctioned. Schools in the top right cell are performing wonderfully. They are doing what we want all schools to do: provide their students with both high proficiency and high growth. For the schools in these two cells, AYP measures accurately reflect their educational outcomes.



Unfortunately, not all schools fall in these two cells. The problems with AYP are clearly evident in the remaining two cells. In the top left are schools whose students are meeting their AYP goals, but where little growth is occurring. Most often found in affluent communities where high-test scores go hand-in-hand with family income, these schools can be referred to as “slide and glide” because they appear to be resting on the laurels of their students. It is important to understand that NCLB does nothing to hold these schools accountable for providing their students with the annual growth to which they are entitled. In a global economy characterized by fierce competition for demanding jobs that pay high salaries and benefits, this is a highly significant shortcoming.

The bottom right cell contains schools with high growth, but low proficiency. These schools have succeeded in academically “stretching” their students, but given how far behind

they were when they entered school, the schools have not yet been able to raise them to proficiency. These schools, while not bringing their students to AYP-required levels, are clearly helping them improve their academic performance, yet still face sanctions under current law.

When NCLB was drafted, some educators advocated the inclusion of growth in the measurement of AYP so schools that helped their students grow would be seen as doing their job, but their proposals were turned down. This has reinforced a conspiracy theory widespread in the K-12 world that explains the exclusion of growth in calculating AYP as part of a plot to make the nation's schools look bad (the high failure rate of AYP) so citizens would give up on public education, thus opening the door for vouchers.

The truth, of course, is very different. The refusal to include growth as an alternative for meeting AYP was a bipartisan decision. Supporters believed that even if students — primarily of minority and low-income backgrounds — grew every year, many would never hit real-world standards by graduation because they had entered school so far behind. Their position was quite principled — “we are not going to abandon these children” — and so they rejected growth measures and insisted on holding schools accountable for getting all their students to proficiency as the sole means of hitting their AYP targets.

Can All Children Achieve High Standards?

Some observers believe that most children who enter school well below grade level will be unable to reach proficiency by graduation. In *Class and Schools*, for example, Richard Rothstein argues that factors beyond the control of schools have such an overwhelming impact on student achievement that only a massive infusion of new funds can overcome the major deficiencies in health, nutrition, socialization and income that hobble these students.

Others are agnostic: maybe all students can, maybe all students can't reach high standards. For now, they argue, schools should be focused on ensuring that each student gets a *year's worth of growth every year* from wherever they start in September. Knowing that schools fall far short of this goal each year — for example, 68 percent of Tennessee K-8 schools did not provide students a year's worth of growth in math in 1996-97 — they conclude that if schools simply did not lose ground, students would be vastly better off. Sustaining cumulative gains over years of schooling would be a dramatic improvement in the status quo. A student entering second grade on grade level in a school averaging 75 percent annual growth would graduate eighth grade at 6.5 years of academic achievement, while one attending a school averaging 140 percent would graduate at 10.4 years of academic achievement.

Still others, building on this idea, argue that an achievable and worthy goal for the nation's schools would be to require them to “stretch” their students beyond a year's worth of growth in a year. That is, not only would they would not lose ground, but they would provide growth at a rate that exceeded predicted performance based on earlier achievement. This accomplishment, they argue, should qualify schools as meeting AYP even if the level of achievement fell short of proficiency.

But because NCLB is the law of the land, this is currently a philosophical discussion. Time will tell whether our schools can in fact bring all their children to proficiency by graduation. But for now let us consider how it is possible to improve AYP without abandoning

the commitment that all children reach proficiency. We call this approach “Growth to Standards” and it is achievable through the introduction of value-added models.

Fixing AYP Without Abandoning Proficiency Through “Growth to Standards”

The essence of the “Growth-to-Standards” approach is to identify schools that are putting their students on growth trajectories to reach proficiency in the future and to credit these schools for that achievement.

Schools would do this by using a value-added methodology that converts the static achievement scores of their students to dynamic growth scores. If students currently performing below their AYP target are on track to reach proficiency by the time they graduate, they would be counted among those meeting their AYP target in the current year. If a school were to place enough of these students on growth-to-standards trajectories, it could meet its AYP goal for the year. Using a growth-to-standards approach, in other words, would reduce the proportion of schools failing AYP, but without abandoning the commitment to proficiency. Several existing approaches — Northwest Evaluation Association’s growth model, Harold Doran’s REACH model and William Sander’s value-added model — could accomplish this.

This approach may be criticized for the same reason that the existing definition of AYP is criticized: it creates what many call a “perverse incentive” for educators to focus like a laser beam on one group of students to the exclusion of all others: those close to but below proficiency. Schools choose to ignore students far below proficiency as well as those whose scores already exceed proficiency, the argument goes, because the prime directive in NCLB is for schools to hit their annual AYP targets.

While this is clearly the logic of the incentive, we do not yet know if this is supported in fact. The growth-to-standards approach described above, like AYP, might simply illuminate the pattern — the gains made by those who start just below proficiency are coming at the expense of those who start the year above it — rather than exacerbate it.

We know this pattern long pre-dates NCLB and has been widespread in poor communities, whether in inner-cities or Appalachia. It explains, for example, the observation made by elementary school teachers that the proportion of precocious students in kindergarten and first grade is sharply reduced by fifth and sixth grades. Faced with so many low performing children, the explanation goes, teachers focus on the bottom of the student distribution so that previous low-achievers get high growth while previous high-achievers get low growth. Sustaining this focus in the early years explains why so few high achieving, low-income children are found in middle school.

When Dr. Sanders applied his growth-to-standards approach to all Tennessee schools in the 2002-03 school year, he learned that 13 percent more schools would meet their federal goals if this alternative means of calculating AYP were accepted by the U.S. Department of Education. But when Sanders looked more closely at its effects — he examined nine Memphis schools all of whose students were minority and low-income (on free and reduced price lunch) — he discovered some troubling results. While some schools met their AYP through the growth-to-standards alternative without denying any of their students adequate yearly growth, others did so apparently at the expense of students who had achieved at higher levels in the past. Seeing no

sense in a trade-off that benefits one group of poor minority kids at the expense of another, Sanders is now at work on a “net” approach: schools would receive credit for students placed on a growth-to-standards trajectory and debits for formerly higher achieving students denied adequate growth in the process.

The Federal Department of Education should encourage states currently using high quality value-added or growth models to conduct pilots over the next few years to determine the impact of a growth-to-standards approach on the measurement of school performance.

Conclusion

It is essential that school reformers understand the limits of what can be done at the federal level so they can concentrate their efforts on the states. The essential focus of NCLB was on bringing up the bottom, itself a significant and long-neglected goal of social justice. It was not and, given political realities, could not have been about ensuring that the bar was raised sufficiently high so that in the future *all* students can graduate as well-educated citizens in an increasingly complex society and as productive workers in the highly competitive global economy of the 21st century.

The political compromises necessary to get NCLB passed left the height of the proficiency bar to the states. Comparing the results for students considered proficient in state tests with those from the National Assessment of Educational Progress makes clear that with a few notable exceptions state standards were set low to begin with. To meet their goals under the new federal law, some states are watering down their proficiency targets. Others have “back-loaded” the time when the highest percentage gains in student performance are to be made in the hope that NCLB will not exist in later years. Still others are sitting by passively allowing schools to subtly and not so subtly encourage their lowest performing students to drop out. And in face of the high failure rates, states are being allowed to adopt “confidence intervals,” a statistical adjustment that further waters down their AYP requirements.

We need, therefore, to create incentives at the federal level to encourage the states to move in different and more desirable directions because it is at the state level that the school system of the 21st century will be created. We’ve argued that value-added assessment could serve as a powerful catalyst for change. Momentum for its use is building. Beyond Tennessee, Pennsylvania, and Ohio, where the Sanders model has been adopted for statewide use, Arkansas, Minnesota, Colorado and Florida have passed legislation calling for the introduction of value-added models. These states have recognized value-added’s many advantages. It traces the academic progress of individual students rather than cohorts. It focuses on ensuring that *all* students, not simply the lowest performers, receive at least a year’s worth of growth in a year. It provides educators with rich diagnostics to improve instruction. And, if allowed, value-added would help schools legitimately meet their AYP goals without abandoning the new federal commitment that poor and minority children graduate from school having met proficiency standards.

But the real prize in having value-added assessment widely adopted is that as the basis for individual-level accountability, it serves as the foundation for the comprehensive reforms, like OPE’s, that will change the organization and governance of K12 schools. Only then will we have in place the means to bring all American students to internationally competitive standards and help meet the nation’s human capital development challenge.