Educational Inequality in the Wake of No Child Left Behind

Adam Gamoran
William T. Grant Foundation

Spencer Foundation Award Lecture
to the
Association for Public Policy and Management

November 7, 2013

Educational Inequality in the Wake of No Child Left Behind

Early in 2002, an extraordinary federal law raised the stakes for high standards in U.S. education. The *No Child Left Behind Act* (NCLB; 2002) required states to set absolute targets for school performance, and to hold schools accountable for meeting those targets. Passed through a rare bipartisan effort, NCLB responded to a variety of aims: for those most concerned with standards and accountability, NCLB tightened the links between performance and evaluation in ways that were unprecedented for federal education policy in the U.S.; and for those who emphasized equity and social justice, NCLB was designed to expose inequalities and to hold schools accountable for the performance of students from all backgrounds (Loveless, 2007; Sadovnik et al., 2007).

Today, the era of NCLB is over. It has fallen victim not to partisan conflict, but to expectations that were impossible to achieve, and to implementation that was incomplete and incoherent. What lessons can we draw from the failure of NCLB? How can future education policies at the federal and state levels address the continuing inequalities in U.S. education, and what role can researchers play in advancing this effort?

In the remarks that follow, I will address these questions by focusing on four key points. First, I will argue that inequality is the central problem of U.S. education, and consequently it is the foremost challenge for researchers and policy makers alike. Second, I will identify the successes and failures of NCLB with respect to reducing inequality, and argue that the successes need not be lost even as the failures are addressed. Third, I will examine the current direction of federal and state education policy in the wake of NCLB's demise, and consider implications for inequality in key areas of post-NCLB reform:

- A focus on achievement growth

- Rigorous and coherent standards and assessments

- Evaluation of teachers as well as schools; and

- Efforts to turn around the achievement trajectories of low-performing schools.

Fourth, I will conclude with a discussion of how policy researchers can address critical gaps in knowledge and, by doing so, may help advance reform efforts in productive ways.

A Focus on Educational Inequality

Headline after headline reports on the mediocre results of U.S. schoolchildren on international assessments of student achievement (e.g., Layton and Brown, 2012). These reports, in my view, miss the main story. The primary problem with U.S. school performance is not that it is too low, but that it is too unequal. For example, on the 2011 Trends in Mathematics and Science Study (TIMMS), a survey of mathematics and science performance in 55 nations, U.S. fourth graders tested in mathematics ranked near the middle of the distribution of countries, comparable to several European nations but far below the leaders such as Singapore, Taiwan, and Japan (Provasnik et al., 2012). Yet when the U.S. sample is restricted to school districts with fewer than 10 percent of students on free and reduced-priced lunch – i.e. districts with few poor students – average scores were equal to those of the top-scoring countries. At the same time, in school districts with 75 percent or more of students on free and reduced-priced lunch – those with the highest concentrations of economically disadvantaged students – average scores were much lower, comparable to Italy and Spain and near the bottom of the ranks of developed countries.

Inequality by Geography

Differences among states constitute another dramatic dimension of unequal educational outcomes. Again drawing on international assessments as a benchmark, high-performing states come close to the top-performing nations, while low-performing states lag far behind. On the 2011 TIMSS, for example, several states participated in the eighth grade mathematics assessment. While Massachusetts scored above Japan and nearly equaled Taiwan and Singapore, Alabama's performance placed it in the bottom quarter of participating countries (Provasnik et al., 2012).

By calibrating the performances of states that participated in TIMMS with the same states' results on the National Assessment of Educational Progress (NAEP), a U.S. national assessment given to a representative sample of students in every state, researchers at the National Center for Education Statistics (NCES) were able to compare the performance of all states to international benchmarks in eighth grade mathematics performance (NCES, 2013; see also Hanushek, Peterson, and Woessmann 2012). Findings from this analysis reinforce the conclusion that differences among states constitute a major axis of inequality in the U.S., as state performance levels varied from those that nearly equaled the world's highest performing countries (e.g., Massachusetts, Vermont, Minnesota, New Jersey, New Hampshire) to those with scores well below the U.S. average and lower than nearly any other western nation (e.g., Mississippi and Alabama).

Inequality in Degree Completion

School achievement is not the only outcome for which middling results obscure wide inequalities. At one time, the U.S. led the world in rates of college completion, but that record has long since been eclipsed, primarily by East Asian countries (OECD, 2010). For example in 2008, about 41 percent of Americans aged 25-34 had obtained college degrees at the associate or

baccalaureate levels. This compared to 63 percent in Korea and 57 percent in Japan and Canada, and ranked sixteenth overall. Yet the average U.S. rate reflects a wide range of rates among states; for example Massachusetts, at 53.4 percent, would have ranked fourth in the world, whereas the five states with rates of college completion below 30 percent New Mexico, West Virginia, Nevada, Louisiana, and Arkansas) are laggards in the world marketplace (Lee and Rawls, 2010).

As with test scores, high school and college completion rates continue to differ dramatically by race/ethnicity and socioeconomic background. For instance, among 25-34 year olds in 2010, 70.7 percent of Asian Americans and 49 percent of Whites had obtained college degrees at the associate's level or higher, compared to 30.3 percent of African Americans and 19.9 percent of Hispanics (Lee and Rawls, 2010). Among young persons born between 1979 and 1982, about 80 percent of those from the top quartile of family income enrolled in a four-year college, whereas the comparable figure for those in the bottom quartile was only 29 percent (Bailey and Dynarski, 2011). This gap of 51 percentage points represents an increase over two decades earlier, when the gap was 38 percentage points.

By these reckonings, the most salient feature of educational performance in the U.S. is its unequal distribution along dimensions of social origins and geography. Reinforcing this view, a recent assessment conducted by the Organization for Economic Cooperation and Development (OECD), a consortium of developed nations, found the U.S. to be one of four countries with "both below-average performance and large social disparities" (OECD 2013a, p. 10; the others were Germany, Italy, and the United Kingdom). On a measure of literacy, educational inequality by the socioeconomic status of students' families was greater in the U.S. than in any other nation (OECD, 2013b).

Inequality is the Problem

The low average performance of U.S. students on international assessments is not merely a reflection of low scores at the bottom of the achievement distribution. On the contrary, the U.S. lags behind the top-scoring nations at every performance level (Hanushek, Peterson, and Woessmann, 2010). Yet it is the prevalence of low performers, more than the dearth of high performers, that is most problematic for economic progress and civil society. Although the U.S. leads the world (among nations participating on international assessments) in the number of low-achieving students, it also leads the world in the number of high performers (Petrilli and Scull, 2011); this occurs in part because the population of the U.S. is large, and in part because the degree of inequality is high. In other words, even though high-achieving U.S. students tend to score below the highest achievers of the top performing nations, there are nonetheless an extraordinarily large *number* of high achievers in the U.S. For this reason, the markers of elite accomplishment in U.S. society – such as winning more Nobel prizes than any other nation (Bruner, 2011; Stephens, 2013), and establishing almost as many patents each year as those from all other nations combined (U.S. Patent and Trademark Office, 2012) – are likely to persist. The U.S. system of higher education continues to be the envy of the world as represented by continuing waves of international student enrollment (Project Atlas, 2012), and our scientific infrastructure is unparalleled (National Science Board, 2012), so the prospects for sustained economic and scientific leadership are strong, despite the pressures of international competition (National Research Council, 2007).

Meanwhile, students who do not achieve even a basic level of academic performance, or do not complete at least a high school education, are limited in their capacity to contribute to the U.S. economy and instead tend to hamper economic progress (Levy and Goldin, 2010). Thus, even though average scores of U.S. students fall below those of their counterparts in the highest-performing nations at every achievement level, it is the prevalence of low achievers rather than the shortfall of high achievers that gives greatest cause for alarm. As Belfeld and Levin (2012, p. 2) explained, "purely from an economic perspective – leaving aside important questions of social equity – opportunity is being lost on a large scale."

The drag on economic progress is not the only reason to be concerned about unequal school performance. Educational inequality is also socially divisive, for at least three reasons. First, as sociologists have long recognized (e.g., Durkheim, [1925] 1973; Parsons, 1959; Dreeben, 1968), education provides a common socializing experience that forges shared bonds despite differences in origins, and as educational differences are exacerbated and are increasingly aligned with variation in origins, the socializing dimension of education becomes less potent. Second, education offers opportunities for social networks that break the boundaries of families and communities, and these networks help knit the fabric of American society; greater disparities in educational opportunities mean fewer cross-cutting ties can be established (Putnam, 2000; Stiglitz, 2012). Third, of course, educational outcomes are predictive of future economic outcomes, so as education becomes increasingly stratified by social origins, the prospects for social mobility across generations are diminished. For these reasons, national and international leaders from big-city mayors to Pope Francis are recognizing that, as President Obama (2013) puts it, inequality is "the defining challenge of our time."

No Child Left Behind as a Response to Educational Inequality

Addressing educational inequality was a major focus of NCLB, as states were not only required to report the proportion of students who reached the proficiency target each year, but to disaggregate those reports on a variety of dimensions including race/ethnicity, free lunch status, disability status, and English language learner status. Hence, one of the major accomplishments of NCLB was to expose inequalities that were previously obscured. Many schools with high test scores on average were forced to recognize internal differences among students from varied backgrounds, and dramatic inequalities between and among schools came to light (Gamoran, 2007a). NCLB also managed to focus the attention of educators on finding new ways to improve teaching and learning: studies of NCLB implementation found teachers and principals seeking new approaches that could help more students succeed on the assessments (Hamilton, Stecher, and Berends, 2005; Stecher et al., 2008). Yet NCLB fell far short of its aim of bringing 100 percent of students to the proficiency target. In 2011, about half the schools in the country failed to make "Adequate Yearly Progress" (the marker of success under NCLB, indicating that a school had met its target for the required proportion of student scoring proficient on state assessments; see Usher, 2011). By 2014 – the NCLB deadline – virtually *all* schools would miss the mark, because few if any schools have reported 100 percent of their students meeting the NCLB target.

Achievement Gains Fell Short of Expectations

The unrealistic nature of expectations under NCLB was easy to recognize, and indeed many writers pointed out the challenges early on (e.g., Linn, 2003; Metz, 2005; Gamoran, 2007a). In talks as early as 2004 I pointed out that "*no* school in the U.S. has *ever* made the degree of progress that the *typical* school will have to make to meet the demands of NCLB" (Gamoran, 2007a). Reading performance in the state of Wisconsin provides an instructive

illustration for test score patterns in the NCLB era. Like many states, Wisconsin took a gradual approach to annual increases in the percent of student required to meet the proficiency target: no progress was required for the first three years, then the standard jumped from 60 to 67 percent proficient, followed by two more years of level performance and then another 7 percentage point jump, and so on until 2010, when the required end goal of 100 percent proficient by 2014 meant that schools had to increase their percent of students proficient by 7 percentage points *every* year (Evers, 2011). In reality, performance was mostly flat throughout the NCLB era; for example in grade 4 reading, about 80 percent scored proficient from 2002 to 2011. As a result, the typical school in Wisconsin met Adequate Yearly Progress from 2002 until 2010, when the proficiency standard began to exceed 80 percent (Evers, 2011).

At the national level, test scores rose modestly during the NCLB era, as represented by NAEP trends in reading and mathematics between 1999 and 2008 (Rampey, Dion, and Donahue, 2009). Moreover, NAEP scores exhibited some evidence of gap-closing, particularly in age 9 reading and age 13 mathematics, in which the progress of African American students was significantly faster than that of white students. Yet the advances fell far short of NCLB's ambitious goals. At current rates of progress in age 13 mathematics, for example, the average White student (not all students, but the *average* student) would score proficient in 2021, and the average African American student would reach the same standard in 2043. Clearly if our nation is serious about addressing educational inequality, more was needed than what NCLB had to offer.

Why Did NCLB Fail?

NCLB fell short of its goals for three reasons. First, as I have already explained, the expectations set for NCLB were unrealistic. The notion that schools would bring 100 percent of

students to meet a rigorous achievement standard in a mere 12-year time frame was beyond the capacity of even the most well-resourced schools and districts, let alone that of schools facing challenging fiscal, social, and academic contexts that confront most educators across the nation. An accountability system in which none can meet the standard is of no value in motivating behavior.

Second, as Harris (2011) has persuasively argued, NCLB violated the cardinal principle of effective evaluation systems: evaluating persons on measures that they can control. Average performance levels of students, the focus of NCLB's evaluation metrics, are not under the control of educators, because student achievement is affected by many other conditions including students' home lives, opportunities in their neighborhoods and communities, and the effectiveness of schooling in previous years (Rothstein, 2004; Gamoran, 2007a; Harris, 2011). By holding schools accountable for the absolute levels of student achievement (as opposed to, say, *increases* in achievement from one year to the next), NCLB held schools to an impossible standard and set them up for failure.

The third reason that NCLB failed to meet its aims was that even the good ideas it offered for how to achieve high and equitable performance were implemented weakly and inconsistently (Gamoran, 2007b). Although its accountability system received most of the attention, NCLB (2002) concerned more than accountability: it also prescribed concrete steps that educators could take to meet accountability demands, such as public school choice, free tutoring for struggling students, placing a "highly qualified teacher" in every classroom, and employing evidence-based instructional methods. As I and several colleagues have recounted elsewhere (Gamoran, 2007b) none of these strategies was effectively deployed. To offer perhaps the most compelling example, one-on-one tutoring is widely recognized as an effective strategy for helping students

catch up academically after they have fallen behind (Elbaunz et al., 2000; Farkas and Durham, 2007). It is an especially effective strategy known for gap closing, because it can be targeted to students who are underperforming and offered for free to those who cannot pay. Drawing on this evidence, NCLB required school districts to make available free tutoring (known as "supplemental educational services," or SES) to low-performing, low-income students in schools that had failed to make adequate yearly progress for three successive years. Despite the promise of this approach, a national evaluation found that it had little impact (Zimmer et al., 2007). The lack of impact may be less surprising when one learns how SES was implemented. As Farkas and Durham (2007) and Burch, Steinberg, and Donovan (2007) have revealed, only about 20 percent of students who were eligible for tutoring actually received it. Moreover, student attendance at tutoring sessions was poor, and group sizes were frequently larger than the 1, 2, or 3 students per tutor found to be effective in the research literature. In contrast to NCLB's tight regulations over teacher qualifications, the qualifications of tutors were not specified, and many states required only a high school degree. NCLB likewise did not decree a system for monitoring tutoring, and few states required it. Finally, although one might have expected that tutoring to help students catch up in class would call for close coordination between tutors and teachers, that rarely occurred (Farkas and Durham, 2007). Under these conditions, it is no wonder that tutoring failed to have the impact under NCLB that it had exhibited in prior research studies. Similar weaknesses have been exposed in other NCLB strategies including school choice, highly qualified teachers, and evidence-based instructional practices (Gamoran, 2007b).

Educational Inequality in the Post-NCLB Era

With the demise of NCLB looming, the administration of President Barack Obama faced a dilemma. Soon, virtually all schools would fail under NCLB's unrealistic standards, and the

administration would be left with a useless law (Dillon, 2011a). Yet the U.S. Congress could not reach agreement on a reauthorization that would set more realistic targets (Dillon, 2011b). Consequently, U.S. Secretary of Education Arne Duncan used his authority under the law to offer waivers from the unattainable provisions of NCLB – the 2014 deadline and the 100 percent proficiency target – to states that would design their own accountability systems to meet specified federal standards, including a focus on achievement *growth* (i.e., gains over time) in addition to absolute test scores, adoption of rigorous standards and assessments, evaluation of teacher performance as well as school performance, and an emphasis on turning around the achievement trajectories of their lowest-performing schools. Thus far, 45 states have applied for waivers and 42 have been granted, along with waivers to the District of Columbia, Puerto Rico, and a consortium of 8 large school districts in California.[1] Will the state plans developed under waiver authority do more than NCLB to reduce inequality in the outcomes of education? On the one hand, the state accountability systems offer achievable aims, unlike NCLB. On the other hand, whereas NCLB included tested strategies to achieve its goals, less is known about the effectiveness of the approach promoted for education reform by the federal government through its waiver criteria. Consequently, the prospects for reducing inequality in this new era of test-based accountability are uncertain.

A Focus on Achievement Growth

A major weakness of NCLB was its insistence on holding schools accountable for absolute levels of student performance, regardless of how students had performed in the previous school year. Although setting the same absolute standard for all students was well intentioned, it was not realistic and meant that schools were accountable for conditions they could not control.

---

[1] The latest information on waivers is posted at: http://www2.ed.gov/policy/elsec/guid/esea-flexibility/index.html

The emphasis on growth – on the progress of individual students over time – is an important corrective to NCLB's misguided approach to accountability. To monitor growth, states have implemented various forms of value-added achievement modeling (ranging from simplistic to sophisticated approaches) to account for prior performance while holding schools accountable for gains over time.

Benefits of focusing on growth. Value-added modeling of achievement trajectories has come under criticisms ranging from the cogent (Braun et al., 2010; Harris, 2011) to the hyperbolic (Collins and Amrein-Beardsley, 2011). The criticisms generally focus on uncertainty associated with attributing test score gains to schools (or teachers – more on that below); the stability of value-added estimates; and the validity of using achievement to measure the effects of schooling. What the critics commonly overlook, however, is the contrast between accountability under NCLB and accountability under the state accountability plans. NCLB failed to distinguish between ineffective schools and schools in which test scores were low because students entered at low levels of performance. Because NCLB held schools accountable for student performance at a given point in time, a school whose students had low test scores when they entered would not have been counted as making "Adequate Yearly Progress," even if individual students were improving. In this way, NCLB mixed up effective and ineffective schools (Linn, 2003). Despite the uncertainties inherent in any rating scheme, accountability systems that rely on value-added calculations to assess whether schools are effective are more accurate than systems that focus on scores at a single point in time. Under NCLB, schools with low-achieving students were sanctioned even though some were helpful to low achievers, and state accountability systems that monitor achievement growth will be much less subject to that error.

The states' focus on achievement growth may also help ensure that students at the low end of the achievement distribution, who are far from NCLB's proficiency criterion, will "count" towards the school's accountability rating. Under NCLB, educators had the greatest incentive to boost the scores of students who were just below the target, and some research suggests that more resources are targeted to these students than elsewhere (Booher-Jennings, 2005). The focus on growth gives credit for improvements among all students, not just those who cross a pre-determined threshold.

Implications for inequality. The transition from an accountability system aimed at absolute scores to one that focuses on both gains and absolute scores may have positive implications for attention to inequality. First, by sorting out effective and ineffective schools more accurately, the state accountability systems will potentially target resources where they are most needed. Second, schools targeted for closure will be those that are actually ineffective, not just those with low average scores. Third, attention to students across the full range of the achievement distribution could result in new resources directed to schools' lowest-performing students, and that too could address inequality concerns in ways that would not have occurred under NCLB.

Rigorous and Coherent Standards and Assessments

Another major change in the post-NCLB era is the adoption of rigorous standards and assessments that are common across a number of states. The Obama administration first promoted rigorous, common standards as a criteria for eligibility for funds under the Race to the Top initiative, a competition among states for funds from the American Recovery and Reinvestment Act (ARRA; 2009). Stimulated by this initiative, 45 states ultimately signed onto the Common Core State Standards. Using additional funds from ARRA, the federal government

funded two large consortia to develop assessments aligned with the Common Core State Standards. Most states that have received waivers from NCLB requirements are committed to using these assessments beginning in 2014 or 2015. Many analysts have concluded that the assessments will be a step up in rigor and coherence compared to the assessments deployed under NCLB (e.g., Carmichael et al., 2010; Schmidt and Houang, 2012). Nonetheless, there remains some uncertainty as to whether "common assessments" will be broadly implemented, as critics on the political right object to "common" and many on the left complain about "assessments" (Rich, 2013). At this point, it seems likely that many states will implement the new assessments in the next school year, but the breadth of implementation remains to be seen.

Research basis for rigorous and coherent standards and assessments. Research on standards-based reform in education provides a strong basis for implementing rigorous, coherent curricula that meet students at their current performance levels and challenge them to advance. At a descriptive level, comparisons across countries and across states within the U.S. demonstrate that those with more rigorous and coherent curricula tend to produce higher achievement among their students (Schmidt, 2011; Schmidt and Houang, 2012). Moreover, studies of curricular upgrading in the U.S. and abroad show that when students encounter more focused and challenging curricular materials, achievement tends to rise (Clune, White, and Patterson, 1989; Gamoran, 1996, 1997, 2000; Gamoran et al., 1997). Both observational and experimental studies demonstrate that students who encounter more demanding opportunities for learning tend to perform at higher levels, and the gains hold for students whose achievement levels are initially low as well as for those whose performance is strong at the outset (Gamoran and Hannigan, 2000; Connor et al., 2009). Evidence of this sort was used successfully by the

framers of the Common Core State Standards to advocate for common standards across states (McDonnell and Weatherford, 2013).

Implications for inequality. Rigorous, coherent, common standards have the potential to support efforts to reduce inequality in at least two ways. First, as noted earlier, differences among states in achievement outcomes constitute one of the most salient sources of inequality in U.S. education. Common standards and assessments will go a long way towards identifying and highlighting this problem. Under NCLB, a state such as Wisconsin could identify as much as 80 percent of students performing at the targeted proficiency level in reading. When the state recalibrated its assessments to meet national standards for proficiency, the rate plunged to about 40 percent – not because students were performing worse, but because the state used a more rigorous benchmark to assess student performance (Defour, 2012). Other states are now undergoing the same reckoning (Hernandez and Baker, 2013). With common assessments, states will no longer be able to hide their students' performance levels behind easy tests or low benchmarks.

The Common Core State Standards not only aim to set rigorous criteria, but also to provide deeper and more coherent curricular aims, which can also be reflected in the common assessments. Because low-income students often face instruction that is fragmented and superficial (e.g., Newmann and Associates, 1996), implementing richer and deeper curriculum in response to the Common Core State Standards could bring special benefits to disadvantaged students.

Evaluation of Teachers as Well as Schools

Perhaps the most controversial aspect of the federal government's criteria for waivers from NCLB requirements is the insistence that student achievement trajectories be used to hold

not only schools but teachers accountable. Under NCLB, accountability rested on schools as organizations, and teachers could be sanctioned indirectly when schools were placed under "improvement plans" after four or more years of failing to make AYP, plans that may have included dismissal of staff. Short of that point, however, teachers did not bear the brunt of NCLB sanctions, and individual teachers were not called out for their own contributions to student test scores. This practice began to change with the Race to the Top competition, as states were required to link students to individual teachers in their longitudinal data systems, and was carried into the waiver provisions as a requirement: to be eligible for a waiver, states had to outline an evaluation system for teachers that would include student test scores as one criterion of performance.

Research on compensation incentives. As with schools, teacher contributions to student achievement trajectories can be assessed with value-added models that aim to isolate the contributions of teachers to student performance (e.g., Harris, 2011). However, value-added analyses of teacher contributions are less robust than those of school effects. First, teachers have far fewer data points (i.e., student test scores) associated in any one year, compared to schools, so value-added estimates are likely to be less reliable, and considerably so in the case of elementary teachers who may have only 25 students or fewer in a year (Harris, 2011). Second, when students are sorted on the basis of test scores (as often happens through the practice of ability grouping), it can be difficult to distinguish effects of teachers from effects of the sorting process, even when value-added models are used (Rothstein, 2009).

Even to the extent that such challenges can be addressed, for example by averaging teacher effects over several years, and ensuring that students are not sorted by ability, it is not clear how the use of test scores to evaluate teachers will contribute to school improvement. A

popular notion about how teacher value-added can be used to drive improvement is that by rewarding effective teachers, all teachers may be motivated to improve. Yet three large-scale experiments with compensation incentives have failed to deliver on this promise: in Nashville, New York, and Chicago, randomized evaluations in which teachers in the "treatment" group had the opportunity to secure substantial bonuses by elevating their students' test scores failed to yield higher achievement among treatment than among control teachers (Springer et al., 2010; Marsh et al., 2011; Glazerman and Seifullah, 2012). In Nashville and Chicago, individual teachers were rewarded, whereas in New York, the financial reward was provided to the school for allocation, yet neither the individual nor the collective approach was effective. These findings suggest that teachers are not primarily motivated by increased earnings, as has been suggested in other studies of teachers' work, and their effectiveness may be more sensitive to working conditions than to compensation (Ingersoll, 2006; Gamoran, 2011).

A small-scale study in one school district uncovered a compensation strategy that seemed to work: instead of rewarding teachers after their students exhibited high test scores, the investigators provided financial rewards at the outset, and notified teachers that they would lose the salary bonuses if the high scores were not sustained (Fryer et al., 2012). This "loss aversion" strategy may have been effective as teachers randomly assigned to this condition produced higher test scores than their control-group counterparts. The generalizability of this approach has yet to be explored.

Another use of teacher evaluations may be to motivate the most effective teachers to work with the most disadvantaged students. The viability of this strategy is uncertain, as more accomplished teachers commonly gravitate towards more advantaged student populations, at least in the case of teachers certified by the National Board for Professional Teaching Standards

(Hakel, Koenig, and Elliott, 2008). Recently, the Talent Transfer Initiative (TTI) attempted to evaluate this policy by offering high performing teachers (i.e., teachers with high value-added scores for student achievement) to move to schools with larger concentrations of disadvantaged students (Glazerman et al., 2013). On the one hand, the TTI study demonstrated that high-performing teachers can maintain their effectiveness with more challenging populations, an important element of any strategy aimed at using teacher evaluations to match effective teachers with low-performing students. On the other hand, a mere 5% of high-performing teachers in the study were willing to make the move to the more challenging environments, a finding that must inject a significant note of caution into policy considerations around this approach.

Implications for inequality. Improving the quality of teaching for disadvantaged students must be a crucial element of any reform package, because students' academic progress is more closely tied to the effects of teaching than any other school-related condition (Rivkin, Hanushek, and Kain, 2005; Hanushek and Rivkin, 2006). NCLB and the state accountability systems offered different strategies to achieve this goal. NCLB demanded that teachers implement evidence-based instructional approaches, yet this feature was never implemented and it is not clear how it could have been, given limited knowledge about what aspects of instruction promote high achievement. As per the federal requirements, the state waivers aim to improve teaching by identifying effective teachers and, through some combination of incentives and sanctions, to elevate the quality of teaching overall, and for disadvantaged students in particular.

While the goal is laudable, it seems clear that a coherent strategy for achieving the goal is lacking. We know what does not work: simply paying teachers more for producing higher test scores. We have also learned that when effective teachers move to schools with lower-performing students, they are able to maintain their effectiveness, but very few teachers seem

willing to make this move. At best, the state waivers have identified a key leverage point –

teacher effects on student learning – but have not identified a viable approach for exerting

pressure on this leverage point to bring about the desired results. It is also possible that teacher

effects on student test scores, in combination with other information, could be used to improve

the pool of teachers by re-training or not retaining the lowest performers. Overall, though, one

must remain skeptical that the emphasis on evaluation of teachers using student test scores will

do much to reduce inequality in student outcomes.

Turning Around Low-Performing Schools

Turning around the achievement trajectories of persistently low-performing schools was

also a requirement for states under NCLB, but the state accountability waivers have taken this

goal in new directions. NCLB prescribed five options for "restructuring" low-performing

schools: replacing the staff, state receivership, private management, converting to a charter, or

"any other major school restructuring." By contrast, federal requirements for state accountability

systems are less prescriptive on the reform strategy, as well they might be since under NCLB,

about 90% of districts selected "other" as their turnaround option in any case (Scott, 2009). The

new state accountability systems identify "priority" schools as the bottom five percent of schools

in test performance, and "focus" schools as the ten percent of schools with the largest

achievement gaps, but do not identify specific strategies that these schools must adopt.

Research on school turnaround.   Limited evidence exists on elements of successful

turnaround plans. Syntheses of studies have identified common elements such as strong

leadership, a committed staff, and a sustained focus on learning and instruction, but the studies

reviewed emphasize practical wisdom and case studies that, while informative, have yet to be

tested with designs that rule out competing explanations for school improvement (Herman et al.,

2008; Learning Point Associates, 2010). One of the strongest studies to date used an "interrupted time series" approach, in which student achievement is monitored before and after a shock to the system, which was in this case a designation of schools as in need of improvement (Bryk et al., 2010). According to this study of Chicago schools, reform conditions that supported improvement included school leadership as the driver for change, the professional capacity of teachers and other staff, positive parent-community ties, a student-centered learning climate, and instructional guidance for teachers focused on the academic demands of the curriculum and teachers' tools to advance student learning. State waiver plans that provide for technical assistance and support for improvement may do well to pursue such research-based practices.

Conversion to charter schools remains a major aspect of many reform plans, yet the evidence on charter schools is mixed. Indeed, the main finding about charter schools is that their effects on student achievement are variable, with some charters outperforming the regular public schools from which their students are drawn, and others performing less well (Zimmer et al., 2009). Evidence about successful charters is growing, particularly based on well-established models such as Knowledge is Power Program (KIPP) Academies (Tuttle et al., 2013). Uncertainty remains, however, about two key aspects of the charter school movement. First, many effective charters seem to derive their success from a cadre of young, energetic staff who work innumerable hours for modest pay, and it is not clear whether the supply of human capital is sufficient to allow for large-scale expansion at the highest level of quality. This concern is reminiscent of the "open schools" movement, which could not be sustained, in part, due to the exhaustion of teachers (Swidler, 1979). Second, the variability in charter school effects may reflect not only differences in the quality of charter schools, but differences in the quality of

schools from which they draw students. As the least effective schools are closed, charter schools that appear highly effective may become more rare.

Implications for inequality. Turning around low-performing schools could do the most of all to reduce achievement disparities, but cases of successful turnaround are exceedingly rare (AIR cite). Impatience with improvement may be a key barrier to success, as studies of comprehensive school reform suggest that improvement may take 3-5 years to take hold (Berends, Bodilly, and Kirby, 2002; Borman et al., 2003; Bryk et al., 2010), yet few reforms remain in place for that duration due to staff turnover and the pressure for quick results. As a result, school turnaround remains a promising strategy whose effectiveness has yet to be demonstrated.

New Directions for Policy Research on Education

Analysis of recent changes in federal education policy demonstrates both the vast scope of the challenge and the paucity of well-tested responses to educational inequality. In this context, policy research can contribute to improving student achievement and reducing gaps by focusing on key challenges and taking advantage of new resources that may afford more powerful research findings.

Focus on Key Challenges

A decade of rigorous research with increasingly careful specification (e.g., Spybrock et al., 2013) has yielded increasing knowledge about programs that may improve outcomes. The What Works Clearinghouse, for example, once derided as the "Nothing Works Clearinghouse," now includes over 700 evaluations including many with positive effects. These advances, however, have not yet translated into substantial reform on a wide scale. A key limitation that constrains implementation of research findings is the emphasis on average treatment effects with

insufficient attention to what works for whom and under what circumstances (Imai and Ratkovic, 2013; Weis, Bloom, and Brock, 2013). Decisions about implementation are not made in the abstract, but in particular contexts for specific participants, and knowledge of variation in treatment effects is needed to make such decisions. This issue is especially important when reducing disparities is the goal, as knowledge of programs that work especially well for disadvantaged students would be the quickest way to close gaps. Yet knowledge of programs that work similarly for individuals of varied backgrounds can also have gap-closing effects if they are applied in the most needy circumstances.

A framework for examining heterogeneous treatment effects. Recently, Weis, Brock, and Bloom (2013) provided a framework for identifying and assessing variation in the effects of programs, both at the individual and the contextual levels. A key notion in their framework is the "treatment contrast," the difference in conditions experienced by members of the treatment and control groups. Often, it is not correct to assume that members of the control group have no exposure to the treatment, or to an experience similar to the treatment. If the treatment is an early childhood education program, for example, some control group members may receive early childhood education experiences aside from the specific treatment being studied, while others may not. If control group members (or control sites in the case of cluster-randomized studies) differ in their experiences of treatment-like conditions, the contrast between treatment and control experiences (the "treatment contrast") is likely to vary, resulting in heterogeneous treatment *effects* across individuals or groups. This insight implies that measuring the treatment experiences of treatment group members, as commonly occurs in experimental studies, is insufficient; to understand the treatment contrast a researcher must examine the experiences of both treatment and control group members. In addition to the treatment contrast, Weis, Bloom,

and Brock (2010) explore ways that differences among participants and among program contexts may introduce variation in treatment effects.

New approaches to studying mechanisms. Another key challenge for policy researchers is to understand the mechanisms through which programs and policies operate. To date, many rigorous studies treat programs as if they were black boxes, seeking a positive or negative judgment without aiming to understand how the outcomes are reached. The next generation of policy research in education will advance if it offers more evidence on mechanisms so that the key elements of programs can be supported, and the key problems in programs that fail to reach their goals can be repaired.

Assessment of mechanisms is no simple matter, as recent writers have demonstrated (Green, Ha, and Bullock, 2010; Hong, 2012; Jo and Stuart, 2012). When treatments are randomized, experimental studies yield unbiased estimates of treatment effects. Yet estimates of mechanisms may still be vulnerable to selection bias if there are omitted variables that predict both the hypothetical mechanism and the outcomes. In such cases, the potential mediator is confounded with the outcome and the causal role of the mediator is ambiguous. Fortunately, recent statistical advances offer new ways to address this challenge, including principal stratification (Page, 2012), marginal mean weighting (Hong and Nomo, 2012), instrumental variables (Raudenbush, Reardon, and Nomi, 2012), and causal mediation analysis (Imai, Keele, and Tingley, 2010). For example, Imai et al. (2011) explain that mediation effects can be identified if one assumes that there are no omitted confounders of the treatment and mediator, nor of the mediator and outcome. This assumption of "sequential ignorability" maintains that omitted variables can be ignored in the sequence from treatment to mediator to outcome. In randomized experiments, the first part of this assumption is satisfied through randomization of

the treatment, and the second part may be satisfied by incorporating pre-treatment confounders of the mediator and outcome into the statistical model. Imai and his colleagues developed causal mediation analysis to estimate mediation effects under this assumption. They also provide a sensitivity analysis to examine the consequences of violating the assumption.

New Resources to Support Policy Research in Education

In addition to statistical advances, important new resources are available that will allow policy researchers to address new questions about educational inequality and to address old questions in more powerful ways than in the past. For example, all states now have longitudinal data systems that link students to their schools and, in most cases, to their teachers, over time, from grades 3 to 8. These data allow a variety of time series analyses that would not previously been possible. Although observational rather than experimental, these data enable strong designs that allow researchers to rule out many types of confounding conditions. Of course, policy research requires information not only on how students respond to their schooling experiences, but on the nature of those experiences. Some of these experiences can be identified through administrative records; for example, class size, charter school attendance, participation in gifted and talented programs, and other structural aspects of schooling are already available in many of the state longitudinal data systems. Other conditions, however, require additional surveys or observations to link students' experiences to their achievement trajectories.

Another new data resource satisfies this need in a novel way. The Measures of Effective Teaching (MET) project collected video data on classroom instruction in about 3,000 mathematics and English language arts classrooms. The main purpose of the study was to identify measures of instruction that correlated with teacher contributions to student achievement (Kane et al., 2013). Identification of measures of teacher effectiveness was an important

contribution in its own right. In addition, however, the MET research team has made its data available for secondary analysis by researchers pursuing a wide range of questions about the relations between students' opportunities for learning and their achievement outcomes over time. These data also constitute a new resource for policy researchers on educational inequality.

Until now, efforts to conduct nation-wide studies of student learning have been constrained by the fact that each state administers its own assessment. Combining outcomes for different states within the same study has been difficult. The new assessments adopted to measure progress towards the Common Core State Standards will facilitate cross-state analyses in ways that were not previously feasible (or required strong assumptions). These data will be especially timely as the nation moves from a single accountability system as demanded by NCLB, to 50 distinct accountability systems as adopted under the NCLB waivers. Comparisons across state policy environments and across specific accountability provisions is a fruitful area for new research that would take advantage of both new data and new policy variation.

Conclusions

Today we commemorate the passing of No Child Left Behind. Although NCLB is effectively dead, efforts to reduce inequality continue. NCLB made important progress, but it was not designed to succeed. Will new approaches represented in the state accountability systems succeed? Implementation was a weak point for NCLB, and if the same occurs with the state accountability systems, the results are unlikely to differ. Enough promising evidence exists, however, to demonstrate that inequality *can* be addressed: not quickly enough to satisfy the unrealistic fantasies of many observers, but enough to make a real difference in the lives of disadvantaged children.

References

American Recovery and Reinvestment Act of 2009, Pub. L. No. 111-5, 123 Stat. 115 (2009).

Available at: http://frwebgate.access.gpo.gov/cgi-

bin/getdoc.cgi?dbname=111_cong_bills&docid=f:h1enr.txt.pdf

Bailey, M. J., and Dynarski, S. M. (2011). Inequality in postsecondary education. Pp. 117-131

in G. Duncan and R. Murnane (Eds.), *Whither opportunity*? New York: Russell Sage

Foundation.

Belfield, C. R., and Levin, H. M. (2012). *The economics of investing in opportunity youth*.

New York: Teachers College, Columbia University. Available at:

http://www.civicenterprises.net/MediaLibrary/Docs/Belfield-

Levin%20Economics%20Investment%20OppYouth%20Sept%202012.pdf

Berends, M., Bodilly, S., and Kirby, S. N. (2002). *Facing the challenges of whole-school reform:*

New American Schools after a decade. Santa Monica, CA: RAND.

Borman, G.D., Hewes, G.M., Overman, L.T., and Brown, S. (2003). Comprehensive school

reform and achievement: A meta-analysis. *Review of Educational Research, 73*, 125-230.

Braun, H., Chudowsky, N., and Koenig, J., Editors. (2010). *Getting value out of value-added:*

Report of a workshop. Committee on Value-Added Methodology for Instructional

Improvement, Program Evaluation, and Accountability, Washington, DC: National

Academies Press. Available at: http://www.nap.edu/catalog/12820.html

Bryk, A.S., Sebring, P.B., Allensworth, E., Luppescu, S., and Easton, J.Q. (2010). *Organizing*

schools for improvement: Lessons from Chicago. Chicago: University of Chicago Press.

Bruner, J. (2011). American leadership in science, measured in Nobel Prizes. *Forbes*, October

5.

Burch, P., Steinberg, M., and Donovan, J. (2007). Supplemental educational services and NCLB: Policy assumptions, market practices, emerging issues. *Educational Evaluation and Policy Analysis, 29*, 115-133.

Carmichael, S. B., Martino, G., Porter-Magee, K., and Wilson, W. S. (2010). *The state of state standards – and the Common Core – in 2010.* Washington, DC: Thomas B. Fordham Institute.

Clune, W. H., White, P. A., and Patterson, J. (1989). *The implementation and effects of high school graduation requirements: First steps towards curricular reform.* New Brunswick, NJ: Consortium for Policy Research on Education.

Collins, C., and Amrein-Beardsley, A. (2011). Review of Harris, D.N., *Value-added measures in education: What every educator needs to know* (Harvard Education Press, 2011). *Education Review,* November 23. Available at: http://edrev.info/reviews/rev1126.pdf

Connor, C. M., Piasta, S. B., Fishman, B., Glasney, S., Schatschneider, C., Crowe, E., Underwood, P., and Morrison, F. J. (2009). Individualizing student instruction precisely: Effects of child by instruction interactions on first graders' literacy development. *Child Development*, *80*, 77-100.

Defour, M. (2012). Less than half of state's students measure proficient under new national standards. *Wisconsin State Journal*, July 17.

Dillon, S. (2011a). Most public schools may miss targets, Education Secretary says. *New York Times*, March 9. Available at: http://www.nytimes.com/2011/03/10/education/10education.html?pagewanted=all&_r=0

Dillon, S. (2011b). Overriding a key education law. *New York Times*, August 8. Available at: http://www.nytimes.com/2011/08/08/education/08educ.html?pagewanted=all

Dreeben, R. (1968). *On what is learned in school*. New York: Addison-Wesley.

Durkheim, E. ([1925] 1973). *Moral education*. New York: Simon & Schuster.

Elbaunz, B., Vaughn, S., Hughes, M. T., and Moody, S. J. (2000). How effective are one-to-one tutoring programs in reading for elementary students at risk for reading failure? A meta-analysis of the intervention research. *Journal of Educational Psychology, 92*, 605-619.

Evers, T. (2011). *An AYP primer: A summary of the NCLB annual performance review*. Madison, WI: Wisconsin Department of Public Instruction. Available at: http://acct.dpi.wi.gov/files/esea/pdf/aypprimer.pdf

Farkas, G., and Durham, R. E. (2007). The role of tutoring in standards-based reform. Pp. 201-228 in A. Gamoran (Ed.), *Standards-based reform and the poverty gap: Lessons for No Child Left Behind*. Washington, DC: Brookings Institution.

Fryer, R. G., Jr., Levitt, S. D., List, J., and Sadoff, S. (2012). *Enhancing the efficacy of teacher incentives through loss aversion: A field experiment*. NBER Working Paper No. 18237. Cambridge, MA: National Bureau of Economic Research.

Gamoran, A. (1996). Curriculum standardization and equality of opportunity in Scottish secondary education, 1984-1990. *Sociology of Education, 29*, 1-21.

Gamoran, A. (1997). Curriculum change as a reform strategy: Lessons from the United States and Scotland. *Teachers College Record, 98*, 608-628.

Gamoran, A. (2000). High standards: A strategy for equalizing opportunities for learning? Pp. 93-126 in R. D. Kahlenberg (Ed.), *A notion at risk: Preserving public education as an engine for social mobility*. New York: The Century Foundation.

Gamoran, Adam. (2007a). School accountability, American style: Dilemmas of high-stakes testing. *Swiss Journal for Educational Science and Research, 29*, 79-94.

Gamoran, A., Editor. (2007b). *Standards-based reform and the poverty gap: Lessons for No Child Left Behind.* Washington, DC: Brookings Institution Press.

Gamoran, A. (2011). Improving teacher quality: Incentives are not enough. Pp. 201-214 in S. Kelly (Ed.), *Assessing teacher quality: Understanding teacher effects on instruction and achievement.* New York: Teachers College Press.

Gamoran, A., Porter, A. C., Smithson, J., and White, P. A.. (1997). Upgrading high school mathematics instruction: Improving learning opportunities for low-income, low-achieving youth. *Educational Evaluation and Policy Analysis, 19*, 325-338.

Glazerman, S., and Seifullah, A. (2012). *An evaluation of the Chicago Teacher Advancement Program (Chicago TAP) after four years: Final report.* Washington, DC: Mathematic Policy Research.

Glazerman, S., Protik, A., Teh, B., Bruch, J., and Max, J. (2013). *Transfer incentives for high-performing teachers: Final results from a multisite experiment* (NCEE 2014-4003). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Goldin, C., and Katz, L. F. (2010). *The race between education and technology.* Cambridge, MA: Belknap Press.

Green, D. P., Ha, S. E., and Bullock, J. G. (2010). Enough already about black box experiments: Studying mediation is more difficult than most scholars suppose. *Annals of the American Academy of Political and Social Sciences, 628*, 200-208.

Hakel, M. D., Koenig, J. A., Elliott, S. W., Editors. (2008). *Assessing accomplished teaching: Advanced-level certification programs*. Committee on Evaluation of Teacher Certification by the National Board for Professional Teaching Standards, National Research Council. Washington, DC: National Academies Press.

Hamilton, L., Stecher, B., and Berends, M. (2005). *Teachers' responses to standards-based accountability*. RAND Working Paper WR-259-EDU. Santa Monica, CA: RAND.

Hanushek, E. A., and Rivkin, S. G. (2006). Teacher quality. Pp. 1052-1078 in E. A. Hanushek and F Welch (Eds.), *Handbook of the economics of education*, volume 2. Amsterdam: North Holland.

Hanushek, E. A., Peterson, P. L., and Woessmann, L. (2012). *Achievement growth: International and U.S. state trends in student performance*. PEPG report no. 12-3. Cambridge, MA: Harvard's Program on Education Policy and Governance.

Hernandez, J. C., and Baker, Al. (2013). A tough new test spurs protests and tears. *New York Times*, April 19, A24.

Harris, D. N. (2011). *Value-added measures in education: What every educator needs to know*. Cambridge, MA: Harvard Education Press.

Hong, G. (2012). Editorial comments. *Journal of Research on Educational Effectiveness, 5*, 213-214.

Hong, G., and Nomi, T. (2012). Weighting methods for assessing policy effects mediated by peer change. *Journal of Research on Educational Effectiveness, 5*, 261-289.

Imai, K., Keele, L., and Tingley, D. (2010). A general approach to causal mediation analysis. *Psychological Methods, 15*, 309-334.

Imai, K., Keele, L., Tingley, D., and Yamamoto, T. (2011). Unpacking the black box of causality: Learning about causal mechanisms from experimental and observational studies.' *American Political Science Review, 105*, 765-789.

Imai, K., and Ratkovic, M. (2013). Estimating treatment effect heterogeneity in randomized program evaluation. *Annals of Applied Statistics, 7*, 443-470.

Ingersoll, R. (2006). *Who controls teachers' work? Power and accountability in America's schools*. Cambridge, MA: Harvard University Press.

Jo, B., and Stuart, E. (2012). Causal interpretations of mediation effects. *Journal of Research on Educational Effectiveness, 5*, 250-253.

Kane, T. J., McCaffrey, D. F., Miller, T., and Staiger, D. O. (2013). *Have we identified effective teachers? Validating measures of effective teaching using random assignment*. Seattle, WA: Bill and Melinda Gates Foundation.

Layton, L., and Brown, E. (2012). U.S. students continue to trail Asian students in math, reading, science. *Washington Post*, December 11. Available at: http://www.washingtonpost.com/local/education/us-students-continue-to-trail-asian-students-in-math-reading-science/2012/12/10/4c95be68-40b9-11e2-ae43-cf491b837f7b_story.html

Lee, Jr., J. M., and Rawls, A. (2010). *The college completion agenda 2010 progress report*. Princeton, NJ: College Board.

Linn, R. L. (2003). Accountability: Responsibility and reasonable expectations. *Educational Researcher, 32* (7), 3-13.

Loveless, T. (2007). The peculiar politics of No Child Left Behind. Pp. 253-287 in A. Gamoran (Ed.), *Standards-based reform and the poverty gap: Lessons for No Child Left Behind*. Washington, DC: Brookings Institution.

Marsh, J. A., Springer, M. G., McCaffrey, D. F., Yuan, K., Epstein, S., Koppich, J., Kalra, N., DiMartino, C., & Peng, A. (2011). *A big apple for educators: New York City's experiment with schoolwide performance bonuses: Final evaluation report*. Santa Monica, CA: RAND. Available at: http://www.rand.org/pubs/monographs/MG1114.html

National Center for Education Statistics. (2013). *U.S. states in a global context: Results from the 2011 NAEP-TIMSS linking study* (NCES 2013–460). Washington, DC: National Center for Education Statistics.

National Research Council. (2007). *Rising above the gathering storm: Energizing and employing America for a brighter economic future*. Washington, DC: The National Academies Press.

National Science Board. (2012). *Science and engineering indicators 2012*. NSB 12-01. Arlington, VA: National Science Foundation.

Newmann, F.M., and Associates. (1996). *Authentic achievement: Restructuring schools for intellectual quality*. San Francisco: Jossey-Bass.

No Child Left Behind Act of 2001, Pub. L. No. 107-110, 115 Stat. 1425 (2002).

Obama, B. (2013, December 4). Remarks by the President on economic mobility. Washington, DC: The White House. Available at: http://www.whitehouse.gov/the-press-office/2013/12/04/remarks-president-economic-mobility

OECD. (2010). *Education at a glance 2010*. OECD. Paris: Organization for Economic Cooperation and Development.

OECD. (2013a). *Skilled for life? Key findings from the Survey of Adult Skills*. Paris: Organization for Economic Cooperation and Development.

OECD. (2013b). *OECD skills outlook 2013: First results from the Survey of Adult Skills*. Paris: Organization for Economic Cooperation and Development.

Page, L. (2012). Principal stratification as a framework for examining mediational processes in experimental settings. *Journal of Research on Educational Effectiveness*, *5*, 215-244.

Parsons, T. (1959). The school class as a social system: Some of its functions in American society. *Harvard Education Review*, *29*, 297-318.

Petrilli, M. J., and Scull, J. (2011). *American achievement in international perspective*. Washington, DC: Thomas Fordham Institute.

Project Atlas. (2012). *Trends and global data 2012.* New York: Center for Academic Mobility Research.

Provasnik, S., Kastberg, D., Ferraro, D., Lemanski, N., Roey, S., and Jenkins, F. (2012). *Highlights from TIMSS 2011: Mathematics and science achievement of U.S. fourth- and eighth-grade students in an international context* (NCES 2013-009). Washington, DC: National Center for Education Statistics.

Putnam, R. D. (2000). *Bowling alone: The collapse and revival of American community*. New York: Simon & Schuster.

Rampey, B. D., Dion, G. S., and Donahue, P. L. (2009). *NAEP 2008 trends in academic progress* (NCES 2009–479). Washington, DC: U.S. Department of Education.

Raudenbush, S. W., Reardon, S. F., and Nomi, T. (2012). Statistical analyses for multisite trials using instrumental variables with random coefficients. *Journal of Research on Educational Effectiveness, 5*, 303-332.

Reardon, S. F. (2011). The widening academic achievement gap between the rich and the poor: New evidence and possible explanations. Pp. 91-115 in G. Duncan and R. Murnane (Eds.), *Whither opportunity*? New York: Russell Sage.

Rich, M. (2013). School standards' debut is rocky, and critics pounce. *New York Times*, August 15, A11.

Rivkin, S., Hanushek, E. A., and Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica*, *73*, 417-458.

Rothstein, J. (2009). Student sorting and bias in value-added estimation: Selection on observables and unobservables. *Education Finance and Policy, 4*, 537–571.

Rothstein, R. (2004). *Class and schools: Using social, economic, and educational reform to close the black-white achievement gap*. New York: Teachers College Press.

Sadovnik, A., O'Day, J., Bohrnstedt, G., and Borman, K. M., Editors. (2007). *No Child Left Behind and the reduction of the achievement gap: Sociological perspectives on federal Education policy*. London: Routledge.

Schmidt, W. H. (2011). STEM reform: Which way to go? Paper presented at the National Research Council Workshop on Successful STEM Education in K-12 Schools. Available at: http://sites.nationalacademies.org/DBASSE/BOSE/DBASSE_080128

Schmidt, W. H., and Houang, R. T. (2012). Curricular coherence and the Common Core State Standards for mathematics. *Educational Researcher*, *41* (8), 294-308.

Scott, C. (2009). *Improving low-performing schools: Lessons from five years of studying restructuring under No Child Left Behind*. Washington, DC: Center on Education Policy. Available at: http://www.cep-

dc.org/document/docWindow.cfm?fuseaction=document.viewDocument&documentid=3 00&documentFormatId=4588

Springer, M., Ballou, D., Hamilton, L., Le, V., Lockwood, J. R., McCaffrey, D. F., Pepper, M., and Stecher, B. M. (2010). *Teacher pay for performance: Experimental evidence from the Projects on Incentives in Teaching*. Nashville, TN: National Center on Performance Incentives at Vanderbilt University.

Stecher, B. M., Epstein, S., Hamilton, L. S., Marsh, J. A., Robyn, A., McCombs, J. S., Russell, J., and Naftel, S. (2008). *Pain and gain: Implementing No Child Left Behind in three states, 2004-2006*. Santa Monica, CA: RAND.

Stephens, B. (2013). Nobels and national greatness. *Wall Street Journal*, October 14.

Swidler, A. (1979). *Organizations without authority: Dilemmas of social control in free schools*. Cambridge, MA: Harvard University Press.

Stiglitz, J. E. (2012). *The price of inequality: How today's divided society endangers our future*. New York: W.W. Norton & Company.

Tuttle, C. C., Gill, B., Gleason, P., Knechtel, V., Nichols-Barrer, I., and Resch, A. (2013). *KIPP middle schools: Impacts on achievement and other outcomes*. Washington, DC: Mathematica Policy Research.

U.S. Patent and Trademark Office. (2012). Patents By country, state, and year – all patent types. Washington, DC: U.S. Patent and Trademark Office. Available at: http://www.uspto.gov/web/offices/ac/ido/oeip/taf/cst_all.htm

Usher, A. (2011). *AYP results for 2010-2011*. Washington, DC: Center on Education Policy. Available at: http://www.cep-dc.org/displayDocument.cfm?DocumentID=386

Weis, M. J., Bloom, H. S., and Brock, T. *A conceptual framework for studying the sources of variation in program effects.* New York: MDRC. Available at: http://www.mdrc.org/publication/conceptual-framework-studying-sources-variation-program-effects

Zimmer, R., B. Gill, P. Razquin, K. Booker, and J.R. Lockwood. (2007). *State and local implementation of the No Child Left Behind Act, Volume I—Title I school choice, supplemental educational services, and student achievement*. Washington, DC: U.S. Department of Education. Available at: www.ed.gov/rschstat/eval/choice/implementation/achievementanalysis.pdf

Zimmer, R., Gill, B., Booker, K., Lavertu, S., Sass, T. R., and Witte, J. (2009). *Charter schools in eight states: Effects on achievement, attainment, integration, and competition.* Santa Monica, CA: RAND.